ARI Research Note 97-11

Feedback for Skill Acquisition: Preliminaries to a Theory of Feedback

Richard A. Schmidt
University of California

Research and Advanced Concepts Office
Michael Drillings, Chief

May 1997

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

University of California

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1997, May	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) 1985-1989

4. TITLE AND SUBTITLE Feedback for Skill Acquisition: Preliminaries to a Theory of Feedback	5a. CONTRACT OR GRANT NUMBER MDA903-85-K-0225
	5b. PROGRAM ELEMENT NUMBER 0601102A
6. AUTHOR(S) Richard A. Schmidt (University of California)	5c. PROJECT NUMBER B74F
	5d. TASK NUMBER 8526
	5e. WORK UNIT NUMBER C71
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Regents of the University of California Department of Psychology University of California, Los Angeles Los Angeles, CA 90024-5600	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-BR 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Research Note 97-11

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

COR: Michael Drillings

14. ABSTRACT *(Maximum 200 words)*:

In training for skills, feedback about skill proficiency--termed knowledge of results (KR) in the laboratory--is critical to efficient learning. But, while various manipulations of KR in acquisitions can provide immediate benefits for performance, these may disappear in retention tests. In several paradigms, we show that (compared to feedback after each trial) making feedback less "useful" by giving it less frequently, or by summarizing or averaging it after several trials, degrades performance in acquisition, but produces superior learning as measured on retention or transfer tests. Preliminaries to a guidance theory are proposed on retention or transfer tests and are proposed to account for these effects. In this view, frequent feedback has various negative effects that degrade retention, such as (a) the encouragement of maladaptive short-term corrections that disrupt response stability, and (b) the blockage of information-processing activities that lead to the learning of error-detection capabilities. Practical implications of these concepts for Army training procedures are discussed.

15. SUBJECT TERMS

Training — Feedback — Learning — Guidance Theory — Retention

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	71	

FEEDBACK FOR SKILL ACQUISITION: PRELIMINARIES TO A THEORY OF FEEDBACK

CONTENTS

Feedback for Skill Acquisition: Preliminaries to a Theory of Feedback

For centuries scientists who have studied the learning of motor skills have known that information about the learner's success in meeting some environmental goal--often referred to as feedback--is critical for efficient learning (see Adams, 1987, for an historical review). This area of feedback for skill acquisition has attracted considerable attention as a result, mainly in the realm of motor behavior, but also to some extent in verbal behavior as well. Many forms of feedback are usually available to learners, making up the intrinsic information that is normally available during or after the production of a movement, such as seeing that I typed the proper letter, or sensing kinesthetically that a dance movement is smooth and effortless. I suspect that the majority of research effort in feedback and learning has been motivated by the desire to understand how these intrinsic feedback processes operate to provide the natural learning enjoyed by animals and humans throughout their lifetimes.

However, this intrinsic feedback is not easily controlled and manipulated in the laboratory, making it difficult to study experimentally. As a result, many investigators have operationalized such feedback in laboratory settings as knowledge of results (KR)--usually defined as augmented, extrinsic (supplementary to intrinsic feedback), verbal(izable), post-response information about the extent to which the movement has met the environmental goal (Schmidt, 1988, Chapter 13). Especially in many real-world tasks, this feedback information can be redundant with the intrinsic information (Fowler & Turvey, 1978), which is difficult to control. As a result, scientists have chosen relatively simple tasks for their research, preventing learners from using this intrinsic information in a variety of ways (e.g., blindfolds, etc.), so that the remaining intrinsic feedback is not particularly meaningful or useful, and extrinsic feedback becomes the most critical variable for learning. Then, extrinsic information is provided in the form of KR, allowing investigators to examine a host of variations of this form of information for learning[1].

FEEDBACK FOR SKILL ACQUISITION

Those interested in training in a variety of practical situations (such as ARI) have had a slightly different motivation for examining feedback, and slightly different methods. Here, feedback is viewed as a training procedure during an acquisition (or practice) phase. The concern is for how such information can be used to maximize learning, to make it more efficient, to provide increased capabilities for performance in long-term retention, or to enhance generalization to other similar actions or situations. As such, the concerns for training have

centered on questions concerning what kinds of information should be provided, and on its scheduling throughout practice. Other kinds of supplementary information are of interest as well, such as the information that a coach or teacher might provide to an athlete or dancer about how the pattern of action could be improved (e.g., "You bent your elbow that time"), hopefully leading to more effective goal achievement on the next attempts; such information is termed knowledge of performance (KP), or kinematic feedback, because it reports information about movement kinematics. Even through those interested in training have had this slightly different motivation for conducing feedback research, the efforts converge with those of scientists interested in intrinsic feedback properties to generate a field of study that examines artificially produced feedback in practice, and how this feedback interacts with the processes of learning.

Numerous reviews of this literature have been conducted during the past few decades (e.g., Adams, 1968, 1971, 1987; Bilodeau, 1966, 1969; Newell, 1974, 1977; Newell & McGinnis, 1985; Salmoni, Schmidt, & Walter, 1984). This literature has been remarkably consistent--almost to the point of being cliché--that during acquisition (i.e., a practice phase), any manipulation of feedback that makes KR more frequent, more immediate, more accurate, more vivid, or informationally useful enhances learning. This generalization has found its way into most of the textbooks in the area, and has had a profound influence in numerous practical settings, such as the organization of training in athletics, music, and the military, and in the design of simulators. Under the belief that more feedback is always better for training, considerable effort has been made to provide as much feedback for learners as possible. This often results in considerable expense (e.g., as in simulators) to duplicate the kinds of information seen in the actual criterion environment.

However, there is good reason to suspect that this generalization about feedback is at best oversimplified, and at worst incorrect. The present report first summarizes the basis of this argument, which was generated originally by our review and reanalysis of the KR literature (Salmoni et al., 1984). This review revealed that certain variations of KR that were highly effective for performance during practice were not very effective for various kinds of retention tests, which we regard as being indicants of learning. More recently, we at UCLA (supported by ARI) and several others have conducted a number of new experiments aimed at these issues surrounding feedback and learning. Together, these data provide considerable difficulty for earlier viewpoints about feedback's role in learning. This report describes these experiments, and emphasizes the difficulties they provide for the earlier theories about feedback and learning. At the same time, these data have suggested several features that seem to be required for a new theory of feedback, and preliminaries to one such view of feedback are sketched here. And,

application of these principles of feedback to ARI settings are suggested.

LEARNING VERSUS PERFORMANCE EFFECTS OF FEEDBACK

As mentioned in the previous section, the literature on KR and skill learning has been strong in showing that, during practice when KR is present and being manipulated, increasing the learner's capability to use feedback in various ways (through increased frequency, precision, decreased latency, etc.) has had beneficial effects on skills, both enhancing performance and increasing the rate of improvement during practice. The early reviewers concluded on the basis of this evidence that these beneficial performance effects were due to learning--usually defined as some underlying, relatively permanent capability for responding. But this conclusion has ignored the possibility--first recognized long ago (Guthrie, 1952; Hull, 1943; Tolman, 1929)--that the changes in performance seen in practice are not necessarily due to learning, but rather may be due to temporary, transient factors which could disappear as soon as the variation in KR is removed. These factors are emphasized especially when performance on a retention test is required, particularly if the task, or the conditions surrounding it, are changed. In addition, both the relatively permanent (learning) and temporary (performance) factors could be operating simultaneously to influence the performance levels observed in acquisition.

According to Salmoni et al. (1984), a problem is that KR researchers, as well as the earlier reviewers of their literature, have generally ignored this learning-performance distinction, assuming that all of the changes seen in acquisition were based on relatively permanent changes in some underlying capability to perform--what is usually termed learning. That is, until recently, no one considered the possibility that feedback in practice could have any other than learning effects, due in large part to the influence that Thorndike's (1927) thinking about feedback has had for researchers [but see Griffith (1931), McGuigan (1959), Trowbridge & Cason (1932) for exceptions].

However, several aspects of feedback may exert their influence only temporarily. One is the well-known motivating (or energizing) effects of feedback (Arps, 1920; Crawley, 1926; Elwell & Grindley, 1938), where subjects who receive feedback exert more effort and report more positive feelings toward the task than do subjects without feedback. Another factor--which we emphasize in the present report--is the informational property of feedback (Elwell & Grindley, 1938; Newell, 1974, 1977; Salmoni et al., 1984), in which feedback informs about the magnitude and direction of error, and thus directs or guides the learner in terms of how to

correct the error on the next trial. In this way, the subject gradually converges (or is guided) toward the goal with practice. This informational property of feedback forms the basis of the theoretical ideas to be discussed later here, in which the guiding effects of feedback produce effective performance in practice when feedback is present, but may actually interfere with processes necessary for more permanent task learning. Until recently, the learning and performance effects of feedback in the earlier literature have generally been confounded.

However, there is more to the problem than the question of whether the effects of feedback manipulations are permanent or not. Researchers have generally not considered the possibility that, relative to some "standard" way of providing feedback, some other variation in feedback which improves performance in acquisition may actually turn out to degrade performance in a retention test. Evidence for this kind of reversal of effect for performance in practice versus retention is provided here. These reversals are especially strong if the retention test involves reduced feedback, as in many real-world training settings where feedback is degraded or withdrawn when the skill is ultimately used. The claim is that this focus on performance (only) in acquisition has systematically misled earlier researchers with respect to the variations in feedback that are effective for learning, at least as learning is assessed on various kinds of retention tests.

Salmoni et al. (1984) argued that, in order to unravel these performance versus learning effects of feedback, various retention (or transfer) tests are needed, in which the feedback conditions are equated. With equated conditions, the temporary factors associated with a particular feedback manipulation are theoretically held constant, and any resulting differences in performance will therefore be due to the variations in the feedback conditions experienced earlier in the acquisition phase. Of particular interest was a retention test in which all subjects are transferred from various feedback conditions in acquisition to a no-feedback condition. In addition to meeting the requirement of providing equated feedback conditions for the retention test, this kind of test was particularly desirable because it would minimize (a) continued improvements in performance during the retention trials and (b) the trial-to-trial adjustments typical of feedback conditions, giving a maximally stable basis for evaluating performance differences. Of course, other kinds of retention tests (e.g., presenting feedback after each trial) are possible, and may even be preferable to the tests without feedback that we advocated.

In our review of this feedback work (Salmoni et al., 1984), we were able to locate a few investigations which have used no-feedback retention tests, and the findings were generally quite surprising. In two lines of work, variations in feedback that degraded performance in acquisition (relative to the "usual" condition with feedback after each trial) were shown to

improve performance in a retention test performed without feedback. One variable was relative frequency of feedback, or the proportion of trials during practice on which feedback is given (Ho & Shea, 1978; Johnson, Wicks, & Ben-Sira 1981), where low (e.g., 10%) relative frequencies of KR in acquisition degraded accuracy relative to a 100% condition, yet resulted in markedly more accurate performance on a delayed retention test without feedback. Another example involved summary KR, in which a summary of a set of trials (e.g., 20) was provided only after that set of trials had been completed (Lavery, 1962; Lavery & Suddon, 1962). During acquisition, summary KR degraded performance relative to a condition in which KR was given after each trial; but on a delayed retention test without feedback, subjects who had received summary KR performed more accurately than those who had received KR after each trial. These results show that at least some variations of feedback which are beneficial for performance in acquisition are detrimental to the development of a long-term capability to perform.

GUIDANCE ANALOGY FOR FEEDBACK--A WORKING HYPOTHESIS

These early results suggested to Salmoni et al. that feedback's strong informational properties could act in a way analogous to the functioning of guidance. In the guidance literature (e.g., Annett, 1969; Armstrong, 1970; Holding, 1965, 1968), various ways of preventing or minimizing errors are used during practice; examples include physically restraining behavior, providing concurrent augmented information about errors, or giving verbal or symbolic information about up-coming movements. Almost by definition, these techniques are effective in improving performance (relative to a no-guidance control condition) when they are present. But, the usual finding is that, when the relative amount learned is evaluated via a retention test without guidance, a group previously guided in acquisition seldom outperforms the no-guidance control condition, and is frequently poorer in performance (e.g., Annett, 1969). A common interpretation is that the guidance during acquisition in some way encourages the development of a memory structure that employs the guidance cues as an integral part of the representation; therefore, performance is poor in retention when guidance cues are removed. A closely related interpretation is that the guidance cues have acted as a kind of "crutch" during practice, allowing effective performance, but essentially preventing the learner from acquiring what is necessary for effective performance when the guidance cues are removed.

Recognizing that the information properties of feedback fit reasonably well within a

definition of guidance--in that it provides extra information about task performance that drives, or guides, the learner to the correct solution of the movement problem--Salmoni et al. amplified earlier suggestions (Annett, 1969; Holding, 1965, 1968; Lintern, 1980) that guidance-like effects of feedback may also operate to degrade task learning under certain situations. That is, while guiding the learner to the correct movement pattern is certainly an important aspect of feedback's function for learning, this guidance from feedback could be overdone, allowing the learner to become dependent on it, and thus actually interfering with processes that would be effective later, especially if feedback were degraded or no longer present. This kind of view provided at least one way for us to understand how certain feedback manipulations that degrade performance in practice (decreased relative frequency, summary KR) might be more effective for learning--at least as learning is measured on various retention tests. The hypothesis was relatively vague, however, about which processes were being degraded or facilitated by the guidance-like effect of frequent feedback.

In the following sections, I outline the results of experiments conducted in our laboratory over the past several years that were directed at this general theoretical idea. A number of new insights into the functioning of feedback, and their potential applications to ARI training settings, are revealed by these studies. These provided strong suggestions about some of the underlying processes associated with feedback and learning, leading to a preliminary statement of a theory for feedback described in a later section. Several of our paradigms in which these questions have been studied are described next, and the major findings from this work bearing on a guidance theory are outlined.

RELATIVE FREQUENCY OF FEEDBACK

One of our paradigms re-opened issues concerning the role of relative and absolute frequency of feedback, begun some years ago by Bilodeau and Bilodeau (1958). Absolute frequency refers to the number of trials in a sequence for which feedback is provided, whereas relative frequency is the proportion of trials for which feedback is provided, or simply the absolute frequency divided by the number of trials in acquisition. Such variables are fundamental for defining the structure of feedback administration in learning settings, and they have naturally received considerable experimental attention (see Salmoni et al., 1984, for a review). But this variation of feedback has been interesting to our group because of the potential relevance to the guidance hypothesis for feedback, which suggested that reduced feedback could degrade performance in acquisition, but enhance learning.

Early Evidence on Relative Frequency

Bilodeau and Bilodeau (1958) manipulated the relative frequency of feedback in simple linear-positioning tasks, holding constant the number of feedback trials at 10 (i.e., absolute frequency), and varying the number of no-feedback trials between the feedback trials to produce four groups with 10%, 25%, 33%, and 100% relative frequency. Whereas this procedure allows the total number of practice trials to covary with relative frequency, it was at the time easy to assume (with supporting evidence) that such simple tasks (blindfolded positioning) do not profit from no-feedback practice (e.g., Bilodeau, Bilodeau, & Schumsky, 1959; Trowbridge & Cason, 1932), and hence the no-feedback trials between the feedback trials were assumed to be "neutral" with respect to the learning of these tasks; as we shall see, this assumption has since proved incorrect.

Bilodeau and Bilodeau examined the performance accuracy on the trials immediately following KR (every trial for a group with 100% relative frequency, every third trial for the group with 33% relative frequency, etc.) during the acquisition phase (only), showing that the performance in acquisition was almost identical for the four conditions. They argued that absolute frequency was an important variable for learning, and that relative frequency was irrelevant. However, no retention tests without feedback were used, so from our perspective it was not clear as to whether these effects of relative frequency were relatively permanent or only temporary in nature.

Johnson et al. (1981) and Ho and Shea (1978), also using simple linear positioning tasks, extended Bilodeau and Bilodeau's (1958) findings by adding no-feedback retention tests. In both experiments, absolute frequency was held constant at 10, and the relative frequency was varied by altering the number of totals trials (and, hence, the number of no-feedback trials provided between the feedback trials), essentially as Bilodeau and Bilodeau had done. When the trials immediately following the feedback presentation were plotted, there were no differences between groups as before. (This is, of course, tantamount to showing that decreased relative frequency depressed performance in acquisition, because performance on, for example, the 10th actual trial was systematically more accurate as the relative frequency increased; see also Winstein & Schmidt, 1990.) However, in no-feedback transfer tests conducted on a subsequent day, the 10% group retained its performance relatively well, whereas the 100% group suffered considerable decrements. Generally, the performance error on the retention tests was directly related to the relative frequency in acquisition, with the 100%

conditions having greatest error, and the 10% conditions having the smallest. And, when viewed across experiments (Ho & Shea vs. Johnson et al.), the differences between groups appeared to increase as the length of the retention interval increased. Clearly, reducing relative frequency (<100%) aided learning and retention, contrary to the earlier findings viewed only in the acquisition phase (e.g., Bilodeau & Bilodeau, 1958). These data raised the interesting question of how relative frequency, generated by the addition of supposedly "neutral" no-feedback trials during practice, could disrupt performance in acquisition, and yet increase learning and retention.

However, as mentioned above, these later experiments confounded the amount of original practice with relative frequency. One relatively uninteresting interpretation of these effects, therefore, is that the groups with lower relative frequency simply had a larger number of practice trials, and it is well known that practice, per se, influences learning and retention. Our initial experiments in this area were designed to unravel the effects of practice and relative frequency, using delayed no-feedback retention tests to evaluate the relatively permanent effects of the feedback manipulations. Four experiments from our laboratory under the ARI project have manipulations of relative frequency across relatively wide ranges, but with the total number of practice trials held constant or at least controlled.

Relative Frequency Effects With Practice Controlled

Ballistic-Timing Tasks

An initial study used relatively simple ballistic-timing tasks, in which the subject had to learn to move a handle and slide in a three-segment action defined by visually presented target zones (Schmidt & Shapiro, 1986, Experiment 1A). In these studies, the subjects attempted to produce a spatially defined pattern, with the major requirement being that the total movement time be as close to 1000 ms as possible. KR about the total movement time was given after different proportions of the trials in acquisition, so that the relative frequency was manipulated between groups while the number of trials was held constant. After an acquisition phase in which relative frequency of feedback was manipulated, subjects received a retention test without feedback after both 10 min and 24 hours, and these tests provided the evidence about the relatively permanent effects of the relative frequency manipulations in the acquisition phase.

Here, two groups received relative frequencies of either 100% or 33%, with the total number of trials being held constant at 102; there was also a 100% relative frequency condition

with one-third the number of trials (i.e., 34), which provided a control for the number of feedback trials presented in the 33% group. The absolute value of the subject's constant error (i.e., |CE|, a measure of directional error or response bias), decreased at a systematically slower rate during the acquisition phase for the 33% condition as compared to the 100% groups, with all groups converging by the end of the acquisition phase. However, there were no reliable differences between groups in the retention tests, although there was a tendency for the 33% condition to have slightly lower |CE|s on the delayed retention test. In any case, there was certainly no evidence that reducing the relative frequency in the acquisition phase degraded learning, at least as it is measured on these no-KR retention tests.

Spatial-Temporal Patterning Tasks

We were concerned about the simplicity of this task with relatively relaxed spatial requirements, and examined these general questions in several more complicated movements. One of these tasks had the subject learn a particular pattern of elbow movement with three reversals that had to be completed in 800 ms, where the spatial-temporal pattern of reversals, and the total movement time, were stated task goals. This pattern is shown in Figure 1. The trace which ends at 800 ms is the goal action, and the subject's attempt to reproduce it (here, the trace ending at 1000 ms) was provided as feedback on a computer terminal after selected trials in the acquisition phase. Overall performance was measured by root-mean-squared (RMS) error between the subject's movement and the template, computed over the 800 ms of the template.

Figure 1 about here

In an early experiment in this series by Winstein (unpublished) using this task, a 100% KR group and a 10% group were contrasted, where the latter was formed by randomly providing 10% of the trials in acquisition with feedback, so that the total number of practice trials was constant. She found that the 10% condition was slightly depressed in performance during the acquisition phase relative to the 100% condition, but that there was no reliable effect on a subsequent retention test without feedback. In a subsequent study, Winstein (1988, Experiment 1; Winstein & Schmidt, 1990, Experiment 1; see also Schmidt, Shapiro, Winstein, Young, & Swinnen, 1987) used a 33% and a 100% relative frequency condition in acquisition. When subjects were transferred to a retention test without feedback, there was a slight tendency for the 33% group to perform more effectively than the 100% group, but the differences were small and not statistically reliable.

These initial experiments, although producing no reliable effects of the relative-frequency conditions during acquisition on either immediate or delayed retention tests, provided a number of interesting suggestions for future work. First, they showed that reducing the relative frequency in acquisition was not necessarily detrimental for learning, as measured on a retention test without feedback. At first glance, this seems to agree with the interpretations of the Bilodeau and Bilodeau (1958) work, and with the notion prevalent in the earlier literature, that relative frequency was not an important variable for motor learning. But, unlike the Bilodeaus' experiment, our experiments manipulated the relative frequency by holding the number of trials constant, so that reduced relative frequencies were obtained by reducing the number of feedback presentations (i.e., absolute frequency). Absolute frequency, according to the earlier views, is all-important for acquisition, and its reduction should have drastic negative effects on learning. In these studies, we found that, even when absolute frequency was reduced by as much as 90% during acquisition, there were no reliable effects on performance in the retention tests.

This finding that reduced absolute frequency did not affect learning not only alerted us to the possibility that the empirical principles of absolute and relative frequency could be incorrect, they also prompted initial speculation about a guidance hypothesis for feedback. Although the evidence was far from strong, there was at least the possibility that lowered relative frequency might operate by reducing the "crutch-like" guidance effects of feedback in acquisition, requiring subjects to engage in additional information-processing activities, the learning from which would benefit performance in retention. Thus, performance on the retention tests was a joint function of two, opposite effects of low relative frequency: (a) a negative effect because the learner has received less guidance in acquisition, and (b) a positive effect because this reduced guidance has fostered other information processing activities which are beneficial in retention. These two factors could have been more or less balancing each other in these experiments.

Searching for an Optimal Relative Frequency for Feedback

A second set of experiments resembled the ones just discussed in several ways, but here we searched for an optimal relative frequency for learning. If there are essentially two opposing factors operating in the learning process as relative frequency is manipulated, then there should be some optimal relative frequency for learning. Relative frequencies that are too small do not provide sufficient guidance for the learner to approximate the proper movement pattern. Relative frequencies that are too large (close to 100%) provide too much guidance, encouraging

an over-reliance on feedback for trial-to-trial adjustments; this reliance on feedback prohibits "extra" information processing activities that are essential for later performance on retention tests. Supposedly, some intermediate value for relative frequency should provide just enough guidance for accurate production of the proper movement patterning during the acquisition phase, but not so much guidance a reliance on feedback is formed which effectively blocks additional information processing activities. Of course, a traditional view of feedback which posits no negative component associated with high relative frequencies in acquisition predicts that the 100% relative frequency will be optimal. This basic difference in prediction between traditional and guidance notions motivated several additional experiments.

Ballistic-Timing Tasks

One of our unpublished experiments broadened the range of relative frequencies used, asking whether some optimal relative frequency could be found. This experiment used essentially the same ballistic-timing task as that discussed above (Schmidt & Shapiro, 1986, Experiment 1A), and manipulated the relative frequency at 100%, 25%, 33%, and 10% in separate groups--the same relative frequencies as used in the earlier study by Bilodeau and Bilodeau (1958)--and tested for learning on retention tests without feedback. Unlike the Bilodeaus' study, though, the number of practice trials in acquisition was held constant across groups, achieved by withholding feedback on different numbers of trials between trials with feedback to produce the different relative-frequency conditions. Reducing relative frequency in acquisition slowed the rate of |CE| reduction in the acquisition phase. However, in the tests of relative amount learned, there were again no differences between groups, with relative frequencies as low as 10% producing performance in retention that was as effective as that for the 100% group, and the intermediate groups having similar performance levels. There was no clear optimal value for relative frequency.

Spatial-Temporal Patterning Tasks

Again, we were concerned that the failure to find optimal relative frequencies could have been related to the relative simplicity of the ballistic-timing task. This general experiment was repeated, this time with the more demanding spatial-temporal patterning task described earlier (see Figure 1). Relative frequencies were varied at 10%, 20%, 30%, 50%, and 100%, holding the total number of trials in the acquisition phase constant. Whereas there was a tendency for

the lower relative-frequency conditions to show somewhat slower rates of improvement in acquisition, there was again only very small, unsystematic differences on the retention-test performances. And, again, no optimal relative frequency was located.

The lack here, and in the simpler ballistic-timing task, of an optimal relative frequency seemed at first glance to provide difficulties for a guidance hypothesis. However, we were again impressed by the fact that even very low relative frequencies (i.e., only 10%) seemed to result in as effective performance on retention as 100% conditions. We began to wonder about the processes involved in practice that would allow so little feedback to produce such effective performance in retention. And this observation, which was by then present in four experiments with two tasks, ran clearly counter to traditional views of feedback which would expect the 100% conditions to be always most effective. Next, we examined the scheduling of the feedback presentations that are predicted by the guidance hypothesis to be most effective in acquisition in acquisition, which presumably maximize feedback's usefulness and minimize its detrimental effects.

Feedback Scheduling With Reduced Relative Frequency

Faded Feedback

The guidance hypothesis predicts that the schedule of feedback delivery across practice will be an important variable for learning. Specifically, the notion is that relative frequency should be large in early practice when the subject is just acquiring the skill, especially when strong guidance to the proper movement pattern would seem to be the learner's primary goal. However, during later practice when this pattern has been reasonably well approximated, and the learner's goal is to establish consistency, the problem is that the learner must not become reliant on feedback, and it is here in later practice that the feedback should be gradually withdrawn, or "faded" to use the term from the reinforcement literature. In terms of relative frequency, the guidance hypothesis predicts that, for a constant number of trials in acquisition, a gradual withdrawal from feedback in later practice--with a relative frequency across the entire practice phase of less than 100%--should be more effective for learning than a condition with 100% relative frequency throughout. Thus, for particular schedules of feedback withdrawal, the guidance hypothesis predicts that providing less feedback will be more effective for learning. Standard viewpoints about learning, on the other hand, would predict the opposite, as the major contribution to learning under such views is the number of feedback trials received in

acquisition, which is of course larger for the 100% condition.

In one of our experiments (see Winstein & Schmidt, 1990), we contrasted two groups, one with 50% relative frequency and another with 100% relative frequency in acquisition. But here the 50% relative frequency condition was different than in earlier studies in this series, in that the relative frequency was 100% in early practice on each day, and was systematically faded toward 0% toward the end of each day, such that the average relative frequency was 50%. Specifically, the first portion of practice on each day had 22 KR trials, following which we administered sets of eight no-KR trials separated by systematically shorter strings of KR trials. These strings of KR trials were, respectively, 7, 4, 3, 2, and 2 trials in length, with the last two KR trials being the final ones in acquisition for that day. This procedure was repeated exactly on the second acquisition day. Thus, for the 50% group, the absolute frequency of KR was 96, while in the 100% group it was 192. The spatial-temporal patterning task, and the method of delivering feedback via an overlay of the subject's performance with the goal on a computer screen (see Figure 1) were as described earlier. Learning was measured on retention tests without feedback, one given 5 min after the end of the second acquisition session, and another 24 hr later.

<u>Figure 2 about here</u>

The results of this study are provided in Figure 2. There was again relatively strong improvement with practice for both conditions in acquisition, but there were nearly no differences between groups on either day, with the 50%-fade condition showing slightly (nonsignificantly) lower RMS errors than the 100% condition. It was somewhat surprising that, even in acquisition when the guiding properties are thought to be strong, the group with 50% feedback performed at least as well as, and perhaps slightly more accurately than, the 100% group. The critical test of learning concerned performance on the retention tests. On the 5-min test, there was a slight tendency for the 50%-fade group to perform somewhat more effectively, but this .6-RMS error difference was not statistically significant. However, for the delayed retention test (24 hr), both groups had regressed considerably in performance relative to their levels on the previous day, but the 100% group regressed considerably more than the 50%-fade group, and the difference between the two groups was now statistically reliable. Thus, as measured on delayed retention tests without feedback, providing fewer KR trials in acquisition, but administering them according to a "fading" procedure, produced improved retention of this skill. And, this effect seemed to increase in size as the retention interval was lengthened, suggested that the KR fading manipulation operated to retard forgetting. This experiment clearly shows the relevance of relative frequency as variable in learning, contrary to

the earlier conclusions of Bilodeau and Bilodeau (1958) that relative frequency was irrelevant.

However, in order to generate this effect, relative frequency had to be systematically reduced across practice. This provides some support for our guidance hypothesis, which predicts that high relative frequencies are important in early practice when the pattern is just being acquired, but that feedback should be gradually withdrawn in later practice to prevent the subject from becoming dependent on it. It is again interesting that this variable had almost no effects during the acquisition phase, but had its major influence in preventing forgetting across the 24-hr retention interval.

Adaptive Feedback Scheduling

This general idea of faded feedback, where information is given often in early practice when the learner needs it most, and is gradually withheld later in practice when the learner might become dependent on it, is somewhat similar to ideas about adaptive practice scheduling. In this work with tracking tasks, the task difficulty is varied across practice as the subject becomes more proficient (Gopher, Williges, & Williges, 1975; Williges & Williges, 1977). After the learner gains proficiency on a simple version and performance scores are improved, the task is made more difficult, which tends to decrease the subject's performance again and presumably speeds progress to higher levels of proficiency. They key feature here is that the task difficulty is altered as a function of the _subject's_ performance, with some people progressing rapidly according to their abilities and past experiences, and others progressing more slowly.

This idea can be applied to the delivery of feedback as well. In early practice, when performance is relatively ineffective, feedback is given often to bring the behavior toward the goal. With continued feedback, performance improves to some criterion for that stage of practice, and then feedback is gradually withheld. Performance should worsen again without feedback, to the point that feedback should again be given to enhance performance. By adjusting the criterion for providing and withdrawing feedback, one can design a practice schedule that balances the need for the guiding properties of feedback against the potential detrimental learning effects of frequent feedback. Many variations of this scheme could be used, and empirical work should reveal a schedule that is generally effective for learning and retention.

In one way, the fading schedule accomplishes this goal, but it is not strictly adaptive because the schedule is set _a priori_, and the frequency of feedback is not sensitive to learners'

individual proficiencies. But in the so-called "bandwidth-KR" paradigm (Lee, White, & Carnahan, 1990; Sherwood, 1983, 1988) this requirement is achieved[2]. In this method, the rule for assignment of feedback is based on a band of correctness about the target. When the learner's performance falls within the band, the learner does not receive any additional information, but he or she understands that the movement was within acceptable limits of correctness (which is a kind of information feedback in itself). If the performance falls outside the band, only then is feedback given. This kind of feedback schedule is, almost accidentally, somewhat adaptive, because when errors become smaller as a natural consequence of practice in acquisition, more and more of the trials fall within the band of correctness, and feedback is faded across practice. And, the rate of this fading is sensitive to the learner's progress. As one might imagine, this bandwidth procedure during acquisition produces more stable behavior than every-trial KR (Sherwood, 1983), but with slightly larger constant errors, as if the subjects do not correct the movements after those trials on which no errors are reported. But later, Sherwood (1987) and Lee et al. (1990) showed that bandwidth feedback, compared to 100% feedback, resulted in more effective retention performance when feedback was withdrawn, both in terms of constant and variable errors. In addition to bolstering our other findings that reduced feedback frequency can enhance retention, these data suggest that adaptive schedules of feedback might become an important goal in structuring practical learning environments.

Specificity, or Similarity, Hypotheses

In several experiments, lowering the relative frequency in acquisition from 100% has led to better retention without feedback. This is clearly counter to more "traditional" viewpoints about feedback and learning (e.g., Adams, 1971; Bilodeau, 1966; Schmidt, 1975; Thorndike, 1927), and supports of an alternative view that emphasizes the "crutch-like" guidance properties of feedback. However, at least one other interpretation, for which considerable evidence exists, can potentially account for these findings. According to these similarity--or specificity--hypotheses (Henry, 1968; Tulving & Thomson, 1971), for a retention test under a given set of conditions, that set of practice conditions that is most similar to the retention-test conditions will maximize performance. Applied to our experiments, it is possible that the low relative frequency conditions in acquisition, with strings of no-feedback trials separated by feedback trials, are simply more similar to conditions in the no-feedback retention test than are the 100% relative frequency conditions that have feedback after each trial. And, faded feedback (Winstein & Schmidt, 1990) is most effective because the reduced relative frequency makes the faded

condition somewhat more similar to the 100% condition when learning is measured on a no-feedback retention test. In both cases, the task and the feedback inherent in it (including the KR) are presumably learned as a kind of whole, and this Gestalt is then disrupted if anything about the context (here, the feedback frequency) is altered at retention, favoring those subjects who learned under identical conditions. If this hypothesis is correct, then it detracts considerably from a rival guidance view, which claims that more is being learned when feedback is withheld.

Relative Frequency Effects in Retention

One way to test this specificity view is to manipulate simultaneously the relative frequency in the retention test and acquisition conditions. A specificity view predicts that, for a given set of retention-test conditions, the acquisition conditions that most closely match them should be most effective. That is, there might be no acquisition condition that is "best" for learning in general, as different conditions might be more or less effective depending on the conditions imposed in retention. On the other hand, the guidance view emphasizes that a particular variation of feedback in acquisition should be most effective for learning, regardless of the retention-test conditions to which the subjects are transferred. That is, there should be no interaction between the conditions in acquisition and the conditions in retention.

With this rationale, Winstein (1988, Experiment 1) manipulated the acquisition conditions and the retention-test conditions in an experiment using the same spatial-temporal patterning movement used in other experiments discussed here (Figures 1 and 2). There were two levels of relative frequency in acquisition (100% and 33%), administered in two 99-trial sessions on sequential days. After a 10-min retention interval on the second test day, she tested for learning in four different retention conditions of 0% (no KR), 33%, 66%, and 100% relative frequency, such that the two acquisition conditions were completely crossed with the retention-test conditions.

During the acquisition (practice) periods on Days 1 and 2, there was a tendency for the 33% group to perform with slightly (not significantly) less error than the 100% group. However, the major concern was performance on the retention tests, where average RMS error for each of the eight retention-test conditions is summarized in Figure 3. Overall, larger relative frequencies in the retention test produced reliably smaller RMS errors--mainly due to the differences between the no-KR groups (0% relative frequency) and the remaining groups in retention--supporting earlier data that KR is strongly guiding when it is present. Furthermore,

considering the performance on the retention test of the two acquisition conditions (33% vs. 100%), there was an overall tendency (nonsignificant here) for the 33% condition to have smaller errors in retention than the 100% conditions. This was again contrary to what one would expect from the more "traditional" viewpoints about feedback and learning, and tends to support other data here that less feedback can aid retention. These effects may have been weakened by the fact that no long-term retention tests were used, as we have seen earlier here that the effects of feedback variations in the acquisition phase are most prominent with longer retention intervals (e.g., Figure 2).

<u>Figure 3 about here</u>

The most important finding, at least from the point of view of the specificity hypothesis, was that there was no significant <u>interaction</u> between conditions in acquisition and the conditions in retention. For each of the retention conditions, the 33% condition in acquisition always had slightly less error in retention than the 100% condition. Under a specificity hypothesis, one should expect to see that the 100% condition in acquisition was most effective for the 100% condition in retention, but this difference was not present. This lack of an interaction provided no support for a specificity view to account for the benefit of lowered relative frequency in learning. Although the evidence is relatively weak and based on nonsignificant results, the implication is that the low relative frequencies in the acquisition phase produced effects which can be thought of as increased memory strength for this task, and that its expression was generally insensitive to the retention conditions under which the task had to be performed.

<u>Fading Effects</u>

With similar rationale to that just discussed, Winstein and Schmidt (1990) repeated the experiment by Winstein (1988) on faded KR described previously (see Figure 2), but with a few changes. The task and conditions in acquisition were exactly as before, with a 100% and a 50%-faded condition, but now the 24-hour retention test was conducted under 100% KR conditions. (We did not give a 10-min test in this experiment.) A specificity view would predict that the 100% group, with acquisition conditions exactly like those on the retention test, would be more effective than the 50%-faded condition whose conditions were different in acquisition and retention. The results are in Figure 4, where the two acquisition days are shown on the left.

<u>Figure 4 about here</u>

Again, there was a small advantage for the 50%-faded group early on the second day, similar to the effect seen in the earlier experiment (Figure 2), but the between-groups differences were not reliable here. Even with half the feedback being provided, the 50%-fade group performed as well as the 100% group in acquisition. However, on the delayed 100% KR retention test, the 50%-faded group retained performance almost perfectly--similar to their high level of retention between Days 1 and 2--whereas the 100% condition regressed 20% or more in error. As a result, on this retention test the 50%-faded group was significantly <u>more</u> accurate than the 100% group. This difference decreased somewhat across practice on the retention test, but it was still present on the second block of 12 trials. Clearly, in spite of the supposed advantage of the similarity in conditions between acquisition and retention for the 100% condition, the 50%-faded condition was still more effective for learning.

However, there may in fact be an effect of similarity after all, as the difference in favor of the 50%-faded condition in Figure 2 (no-KR retention test, where the similarity effect worked in favor of the 50%-faded group) was 2.1 RMS units; this difference was only 1.7 RMS units in Figure 4, where the specificity effect worked in favor of the 100% group. These are between-experiment effects, however, and so they must be viewed cautiously. But even if taken seriously, the data suggest that the similarity effects are extremely small, at best.

<u>No-KR Acquisition Conditions</u>

Finally, the specificity view predicts a result which almost surely cannot hold. If the conditions of retention have no feedback at all (0% relative frequency), then the most effective condition in acquisition should be a no-KR condition. In tasks of the type discussed here, where learners cannot obtain information about their own errors without external feedback, no-KR conditions generally show no improvement across trials (e.g., Bilodeau et al., 1959), and very poor performance as measured on a no-KR retention test (Trowbridge & Cason, 1932). Overall, then, while there may be a sense in which the similarity per se of the feedback conditions in acquisition and retention are a factor in retention-test performance, there is clearly a limit to such a similarity view in work with feedback (i.e., no-feedback acquisition conditions). Clearly, a more comprehensive account of acquisition, one of which being the guidance hypothesis that considers the processes involved in these low-relative-frequency conditions, is necessary to account for the available facts.

The Learning of Generalized Motor Programs

We have also studied the effect of reduced feedback in a completely different paradigm that examined the acquisition of classes of actions, thought to be controlled by generalized motor programs (Schmidt, 1988). Generalized motor program theory assumes that rapid movements are produced with an invariant underlying temporal structure (its relative timing), and that variations in total (overall) time can be produced by applying parameters to the movement at execution, altering absolute time but maintaining the relative timing. This accounts for the well known phenomenon that the overall speed of an action can be easily increased or decreased, with the temporal structure remaining invariant (Schmidt, 1975, 1985; but see Gentner, 1987). Novel actions with the same relative timing, but with "new" total times, are generated simply by altering a single movement parameter. Tests of this view can be made by in transfer by examining the accuracy of the relative timing structure, either on a previously practiced task version or on a novel version with a different overall time.

Wulf and Schmidt (1989) used this paradigm in an attempt to extend the reduced-relative frequency effects to the learning of classes of actions. If reducing relative frequency aids learning of single tasks as I have discussed in the previous sections, then reducing the relative frequency applied to a group of movements with the same relative timing should increase the learning of the relative timing structure that underlies all of its members. Subjects practiced arm movements to three target-buttons, such that the goal durations of the three segments were in the ratio of 2:4:3, defining the relative timing of the task-class. Feedback was given as errors in the actual segment durations in milliseconds. There were three versions of this task with different overall times (400, 500, and 600 ms in duration), but all of these versions had the same relative timing (2:4:3).

In one experiment, we manipulated the relative frequency of feedback, with one group having feedback on only two-thirds of the trials (divided equally among the three versions); this relative frequency was faded across practice from 100% at the beginning of practice to 50% at the end, for an average relative frequency of 67%. Another condition had feedback on every trial. In immediate and delayed transfer tests without KR, we tested performance on a task where the overall time was novel, but where the relative timing was the same as in acquisition. In Figure 5 are the errors in relative timing for this transfer test, where it is clear that the 67% group had far more accurate temporal structures on this novel transfer test than the 100% group. Reducing relative frequency aided the underlying temporal structure of a class of tasks, as shown on a novel version not practiced earlier.

Figure 5 about here

In a second experiment, Wulf and Schmidt (1989) used the same paradigm again. This time one group never received feedback for the task version with the intermediate overall movement time, with the other two versions receiving 100% feedback, so that the overall relative frequency was 67%. A second group received feedback after every trial. We tested for retention of the task version with the intermediate overall movement time, again with the same relative timing. Note that, in acquisition the 67% group never had feedback on this version, and the 100% group always did, and retention was based on performance in this version. Figure 6 shows the relative timing errors for this intermediate version in acquisition and in retention. During acquisition, the 100% group had more accurate performance on this version than the 67% group, showing the guidance-like properties of feedback for performance when it is present. However, this 67% group, without any feedback on the intermediate version, still improved considerably across practice. In no-KR retention, the 67% group was much more accurate in relative timing on this intermediate version, even though subjects had never received any feedback about this version in acquisition.

Figure 6 about here

Overall, these experiments show that the idea of reduced relative frequency to aid learning of single tasks can be extended to the acquisition of a class of tasks. These data show that reducing relative frequency--whether this reduction be applied to all of the members of the class or to only a single one of them--aids learning of the underlying structure of the entire class. Thus, when retention is tested, either to a version practiced earlier, or to a novel version, the underlying temporal structure is more stable and accurate for the subjects who received less feedback in acquisition. In addition, these data contribute strongly to the already convincing set of experiments supporting the existence of generalized motor programs with invariant relative timing, and provide new direction to the area by showing that relative timing can be learned more effectively by reducing feedback (see also Heuer & Schmidt, 1988).

Overview of Relative Frequency Effects

These studies have unraveled the effects of amount of practice and relative frequency which were confounded in the earlier studies (e.g., Bilodeau & Bilodeau, 1958; Ho & Shea, 1978; Johnson et al., 1981), allowing the role of relative frequency to be viewed over and above the effects of practice trials per se. Reduced relative frequencies--sometimes as low as 20%--produced at least the same learning (measured by retention-test performance) as 100% relative

frequency. Although these are null effects, they are nevertheless quite surprising, as virtually every theory of movement learning that deals at all with feedback as a variable predicts that fewer KRs should degrade performance and learning (e.g., Adams, 1971; Bilodeau, 1966; Schmidt, 1975; Thorndike, 1927). Of course, in these experiments, relative frequency and absolute frequency have been confounded, with the total number of trials being held constant. However, this confound actually makes these experiments somewhat more powerful, as traditional views predict the opposite effect of absolute frequency. And, we show that fewer KR presentations--but when presented in a faded paradigm where relative frequency is systematically decreased across practice--produced more learning than 100% KR. Finally, the data cannot be explained by a specificity view, because the advantage of reduced relative frequency was present regardless of the feedback conditions in retention.

Several competing processes seem to occur in relative-frequency manipulations. First, there is a detrimental effect of lowered relative frequency because subjects need feedback to achieve the proper patterns in these tasks, and to maintain them in later practice. But, at the same time, we conceptualize a beneficial effect of lowered relative frequency, as it prevents a dependency on the guiding properties of KR, and may force the learner into information-processing modes which encourage the development of independent capabilities to perform. Presumably, these two opposing sets of processes more or less balance each other when relative frequency is manipulated. But, when feedback is reduced and faded, the detrimental effects of reduced information are overshadowed by the beneficial effect of a decreased reliance on feedback in later practice, and the faded conditions are superior for learning.

SUMMARY KR

Over two decades ago, Lavery (1962) investigated learning in a paradigm in which the errors on each of a set of trials (e.g., 20) were placed on a graph shown to the learner only after the last trial in the set had been completed. In this method (termed summary KR[3]), various trials intervened between a given trial and the feedback that the subject received about it via the graph (except for the last trial in the set, of course). This method resembles in some ways the so-called trials-delay procedure studied by Bilodeau (1956; Lavery & Suddon, 1962), in which feedback about each trial is presented after a given number of other trials, so that a fixed number of trials intervene between a given trial and its feedback. In both methods, feedback seems to be very difficult to use because of the temporal separation of a given trial and its KR, coupled with the potential confusion of information about a given trial with the performance of some

other. Indeed, relative to the more usual provision of feedback after each trial, summary KR and trials-delay procedures are both strongly detrimental for performance during acquisition when feedback is present, both slowing the "rate" of acquisition and degrading performance at the asymptote.

Lavery's Experiments on Summary KR

From the learning-performance perspective (Salmoni et al., 1984), these variables should be examined on some form or retention test in order to reveal any differential effects on learning. This was done in Lavery's (1962, 1964) pioneering work, and the results [which were largely ignored at the time--see Bilodeau's (1966) review] have provided considerable direction to our own work almost three decades later. Lavery (1962) used three conditions in the acquisition of simple motor tasks: (a) Summary KR presented after a set of 20 trials; (b) Immediate KR presented after each trial; and (c) Both, where the summary was given in addition to the immediate KR. The average percentage correct measures in acquisition and retention are shown in Figure 7. In the acquisition phase, summary KR degraded performance with respect to the Immediate and Both conditions. But in no-feedback retention tests on Days 7-10, the Summary group retained their performance level essentially perfectly, whereas the Immediate and Both groups suffered considerable retention losses, to the point that the Summary group was <u>more</u> accurate than Immediate and Both. That is, relative to the Immediate and Both conditions, Summary KR caused decrements in performance in acquisition, but <u>increased</u> learning, at least as it is measured on the delayed retention tests without feedback. These findings, as well as findings from several other paradigms, figured strongly in our guidance hypothesis. Summary KR, lacking strong guiding properties, might be much less effective for performance in acquisition, but it would prevent dependency-like effects, leading to more learning as evidenced when feedback is removed in a retention test.

<u>Figure 7 about here</u>

One might be tempted to ask why summary KR is so effective for retention. However, another examination of Figure 7 reveals that summary KR might not be so effective, but rather that every-trial feedback is detrimental for learning. This view is supported by contrasts with the Both condition. If summary KR were contributing some information over and above every-trial feedback (information about error trends over the 20 trials summarized, for example), then the Both condition, which also had this information, should also benefit from it in the same way. But notice that the Both condition performed similarly to the Immediate condition in both

acquisition and retention, as if the every-trial KR's powerful guiding properties--and the simultaneous detrimental learning processes--had dominated. To us, this was one of the most provocative effects in Lavery's experiment. Certainly, the proposal that every-trial feedback could be detrimental to learning was a radical proposal given the standard viewpoints about feedback and learning.

Motivated by Lavery's important (though largely ignored) discoveries, we examined these effects further using more complex tasks, more practice, and with some alterations in experimental design. We began to search for an optimal number of trials to be described in the summary KR reports, here termed the summary-KR length. According to the guidance hypothesis, summary KRs that are too short are detrimental to learning because they provide too much guidance and may block important information processing activities. Summary KRs that are too long encourage processing of movement information, but do not provide sufficient direction for the reduction of error. Thus, the guidance hypothesis predicts an optimal summary-KR length, where the negative and positive effects of KR are balanced. We manipulated the summary-KR lengths in acquisition across a wide range, hopefully to reveal this optimal summary length for learning.

Searching for an Optimal Summary-KR Length

Ballistic-Timing Task

Schmidt, Young, Swinnen, and Shapiro (1989) used a variation of a ballistic-timing task, in which the subject moved a slide along a trackway, making two reversals at target zones, so that the overall movement time (MT) was as close to 1000 ms as possible. Four groups of subjects received summary reports of either 1 (essentially immediate KR), 5, 10, or 15 trials throughout an acquisition phase of 90 trials. After each trial, the experimenter plotted the subject's movement time constant error (with respect to sign, e.g., -124 ms) for that trial on a graph of performance against trials. This graph was shown to subjects only after completion of the appropriate number of trials for that summary-KR condition. Previous summary reports were erased, so that only the most recent one was available. Except for the 1-trial summaries, the data points were connected by line segments. No verbal KR was ever given. Retention tests without KR were given after 10 min or 2 days.

The averaged absolute constant errors, |CE|, for each of the treatment groups over blocks of 10 trials in acquisition and retention are shown in Figure 8. In acquisition, increasing the

summary length from one to 15 trials systematically increased errors in acquisition, evidenced both as a somewhat larger asymptote near the end of practice and as a slower "rate" of improvement across blocks. It is clear that summary KR interfered with performance during the learning of these simple motor tasks. In the 10-min retention test, there were no important differences between groups. But in the 2-day retention test, all groups (with the possible exception of the 15-trial group) showed some retention loss across retention interval. However, the 1-trial condition appeared to increase their errors the most, followed in order by the 5-, 10-, and 15-trial conditions, to the point that there was a strong inverse relationship between the summary length in acquisition and errors in delayed retention. No effects were shown with variable errors, VE, and thus the major effects of summary KR were in terms of response biasing. The ordering of groups for retention was, of course, just opposite to that shown during the acquisition phase when KR was being manipulated.

<u>Figure 8 about here</u>

These data showed that summary KR was a variable that depressed performance in direct relation to the summary KR length in acquisition, but appeared to enhance learning as measured on a delayed retention test. This repeats Lavery's general observation very well--but for different summary lengths in a somewhat more complex movement task--and extends Lavery's findings by showing this positive relationship between summary length in acquisition and performance capabilities in delayed no-KR retention. The benefit of the longer summary-KR conditions in acquisition appeared to be in terms of less drift away from the target time, as if the subjects in the shorter summary-KR conditions had lost the capability to detect their own errors. However, we clearly failed to find an <u>optimal</u> summary-KR length for this task; such an optimum, if it exists, lies at or beyond the 15-trial summaries used here, perhaps nearer the 20-trial summaries Lavery used.

<u>Coincident-Timing Task</u>

In a subsequent experiment (Schmidt, Lange, & Young, 1990), we continued to search for an optimal summary KR length as predicted by the guidance hypothesis. We reasoned that an optimum might not have been found in the previous study because of the relative simplicity of the task, where relatively long summaries provided sufficient information to correct errors in movement patterning in the early stages of practice. However, such long summaries seemed less likely to be effective for more complex tasks, where presumably more feedback would be required to generate the appropriate movement patterns, and with less likelihood of subjects

becoming dependent on it. This experiment used the same design as the previous one, with summary lengths of 1 (essentially KR after each trial), 5, 10, and 15 trials during a 90-trial acquisition phase, and no-KR retention tests after 10 min and two days. However, we used a considerably more complex coincident-timing task described earlier (see Schmidt & Young, 1991, for a more complete description). Subjects intercepted a simulated moving object with a hand-operated lever with a back- and forward-swing movement, more or less like hitting a ball with a bat, maximizing a score which was analogous to the distance that the ball would have been propelled. KR was provided as a plot of the score against the trials in the set described by the summary, and was presented to the subject after the set was completed. No verbal KR, nor information about the components of the score (velocity and spatial error), were ever given.

Figure 9 about here

Figure 9 contains the overall performance scores, averaged across blocks of 15 trials in the acquisition phase and retention phases. In acquisition, increased summary-KR lengths generally interfered with performance, with the groups being ordered uniformly according to summary length across the entire acquisition phase. However, this trend was altered in the retention tests, with the 5-trial condition performing slightly more effectively than the 1-trial condition, both of which far outperformed the 10- and 15-trial conditions. This inverted-U effect was largely determined by the 5-trial condition having larger velocities, with all groups having essentially the same spatial errors (both CEs and VEs). Thus, with this task, which is arguably somewhat more "complex" than the ballistic-timing task in the previous experiment[4], we provide evidence for an inverted-U effect of summary-KR length for learning, with the optimum being shorter than for the "simpler" task in Figure 8.

Evaluation of Subjective-Estimation Capabilities

We had suspected that the enhanced learning of various summary-KR conditions occurred because the summaries encouraged the development of stronger error-detection capabilities (Salmoni et al., 1984). The argument is that frequent feedback discourages evaluation of response-produced feedback, and less effective learning of the capabilities for error detection. This hypothesis was studied in the experiment just mentioned (Schmidt et al., 1990, Experiment 1) by giving an additional transfer test after the final retention test on Day 2. In this test, which was not announced until just before it began, subjects were asked to perform an additional 30 no-KR trials as before, but in addition to estimate their score after each trial. We measured the subjects' accuracy in subjective estimation via the within-subject correlation

(across all 30 trials, converted to Z') between objective (actual) and subjective (estimated) scores, which should be sensitive to subjects' capability to detect their own errors (e.g., Schmidt & White, 1972). The highest correlation between objective and subjective errors was for the 5-trial summary condition (.43), followed by the 1-, 10-, and 15-trial conditions in that order (.34, .24, and .21, respectively), indicating that the 5-trial subjects tended to be somewhat more sensitive to their own errors than the others. These correlations were all relatively low, however (as compared to correlations of .90 or more in Schmidt & White, 1972), indicating that none of the groups had developed particularly sensitive error-detection capacities.

We also measured the mean absolute difference between objective and subjective error (e.g., Newell, 1974) as a second estimate of error-detection accuracy. This measure also showed the 5-trial condition to be most accurate (159 score units), followed again by the 1-, 10- and 15-trial conditions (182, 236, and 279 score units). Overall, although the evidence was not particularly strong, there was some support for the notion that the summary-KR conditions generated differential capabilities for subjects to detect their own errors without KR, as the 5-trial summary condition was not only the most effective in overall performance, but tended to have stronger estimated error-detection capabilities as well.

Specificity, or Similarity, Effects

We have interpreted these experiments as showing enhanced learning generated by the strings of no-KR trials before the summary is provided. However, these effects can be at least potentially explained by the notion that the conditions with longer summaries were more similar to the no-KR retention tests than was the 1-trial summary condition, so that improved no-KR retention performance may only be a function of the degree of match between the acquisition and retention conditions. This view is contradicted by the data in Figure 9, in that a specificity view would predict that the 10- and 15-trial summary-KR conditions should have been the most effective for no-KR retention, while we see that these conditions are very ineffective indeed. Nevertheless, we (Schmidt et al., 1990) tested this view more formally in an experiment using the coincident-timing task, with the 1-trial and 5-trial summary conditions being used in acquisition, but where the 24-hour retention-test conditions had KR after each trial (i.e., the 1-trial condition). Here, the specificity view predicts that, compared to the 5-trial summary condition which was nearly optimal in the previous experiment (Figure 9), the 1-trial summary condition in acquisition should be more effective for the 1-trial summary retention test.

Figure 10 about here

The average scores for the two groups in acquisition and retention are shown in Figure 10. As before (Figure 9), relative to the 1-trial condition the 5-trial summary condition appeared to depress performance slightly during acquisition. In the 30-trial retention test (with 1-trial summary KR), overall the 5-trial condition had slightly higher scores than the 1-trial condition, but these differences were not reliable. This provided no support for the prediction from the specificity view that the 1-trial condition should be most effective here. However, when the 30-trial retention test performance was separated into three blocks of 10 trials, an interesting interaction between acquisition conditions and blocks emerged. In Figure 10 the two groups were not significantly different on the first block of trials. But by the third block, the 5-trial group increased performance markedly, whereas the 1-trial condition showed no improvement at all, to the point that the 5-trial group was more effective by the end of the retention test. One interpretation is that the 5-trial condition in acquisition generated an increased capability to use KR to improve performance or to continue learning when KR was subsequently presented in the retention test. These learning effects are somewhat different than the usual ones we see, where the benefits of some condition in acquisition are immediately noticed when conditions are equated in a delayed retention test. In this case, the effects resembled "latent learning" effects seen in verbal learning paradigms, where benefits are not seen until some additional experience has been received. These notions are very speculative at present, and deserve additional attention. Whatever the cause of these effects, this experiment provides no support for the specificity view.

KINEMATIC FEEDBACK

A major concern for many is that KR--defined as information about the achievement of some environmental goal (e.g., whether or not a target was struck)--is not usually the type of feedback used by teachers in music, instructors in the military, or coaches in sport. The problem is that the KR in many real-world tasks is essentially redundant with the intrinsic (i.e., inherent) feedback the learner receives under the usual conditions of performing the action. Thus, if a gymnast produced an error, one would be surprised to hear "You fell on your back that time" as KR, as this information is probably obvious. Rather, what learners seem to require is information about the pattern of action that led to the error so that the movement can be corrected on the next try, such as "You tucked too soon that time." This form of information about movement patterns has been termed knowledge of performance (KP; Gentile, 1972), but

I use the more descriptive term "kinematic feedback" here.

Young's (1988; Young & Schmidt, 1991) experiments suggest that some of the principles of relative frequency, discussed in the previous sections, may be generalized to include kinematic feedback. He used the coincident-timing task discussed earlier, a relatively "complex" laboratory analog of hitting a moving ball with a bat (see also Young & Schmidt, 1990), and gave all subjects KR about the achievement of the movement's goal (i.e., the score, or how far the object traveled in the analogy) on every trial in acquisition and retention. However, during the acquisition phase, four treatment groups were given additional information about various aspects of their movement patterns (kinematic feedback), with a fifth group receiving only KR. Here, the kinematic information referred to the amplitude of the backswing in relation to a goal position of 165^{o} (this position had been shown in our previous work to produce the most effective performance). In the analogy with real-world tasks, receiving KR on every trial is similar to being able to see, in a batting task, how far the ball traveled, whereas kinematic feedback is information that might be supplied by a coach about some aspects of the movement patterning.

Average and Faded Kinematic Feedback

Several forms of this kinematic feedback were given in Young's Experiment 2. One group received this information after each trial during acquisition (Single-Trial). Two other groups received average kinematic feedback, where information about the average backswing position, computed over the previous five trials, was given after each five-trial set in acquisition, and no information about any one particular trial was given. One of these conditions (Average) had feedback uniformly after every five trials, whereas another condition (Average-Fade) had these average-feedback reports faded during acquisition by increasing the number of no-kinematic-feedback trials between reports in the later stages of practice, much as we did in the relative frequency paradigm discussed earlier. Then, all subjects received a retention test without kinematic feedback, but where KR (about movement outcome) continued to be given after every trial.

This experiment not only extends the study of KR variables to kinematic feedback paradigms, but it also provides a basis for additional tests of the guidance hypothesis mentioned earlier. That is, with trials held constant, subjects in the Single-Trial condition received kinematic feedback on each of 200 acquisition trials, whereas the Average condition received only one-fifth the number of feedback reports (40), and the Average-Fade condition received

half that number (i.e., 20) across practice. We asked whether providing fewer kinematic feedback reports would benefit learning in a way analogous to the reduced KR frequency in our other experiments reported earlier here (Winstein, 1988; Winstein & Schmidt, 1990). Therefore, this experiment asked whether the guidance hypothesis might also apply to kinematic feedback.

The average performance for these three groups is shown in Figure 11. During acquisition, performances for these conditions diverged after the 100th trial, with the Average and Average-Fade conditions performing more effectively than the Single-Trial group. These effects persisted, and even increased slightly, into a 1-day and a 1-week retention test. Thus, the Average group, with far fewer feedback reports than the Single-Trial group, learned more as measured on the retention test. And, the Average-Fade condition, with only half the feedback reports as the Average condition, showed almost identical learning, and even some tendency to perform and learn more effectively. This is perhaps more remarkable when one realizes that the period of practice over which the kinematic feedback was reduced most strongly (i.e., Blocks 6-10) was that in which the Average-Fade condition performed the most effectively.

<u>Figure 11 about here</u>

Overall, these data suggest that giving information as averages of the previous group of five trials was better for learning than providing it after every trial. One variable operating here is the amount of feedback provided, the Average conditions having far fewer feedback reports as compared to the Single-Trial condition. It is therefore tempting to suggest that this might be in part a relative-frequency effect. This interpretation is also appropriate to the summary-feedback effects discussed earlier, where summaries of the previous trials after a block of trials (but where each of the trials was described) were more effective for learning than providing feedback after each trial.

However, it is also possible that the summary feedback operates by modifying both the information received and the structure of the practice trials. The summary graph gives information about the average level of proficiency after the block has been completed (as well as other kinds of information, such as variability, trends of improvement, etc.). This form of information may tend to stabilize performance, as the averaged feedback is a more stable measure of patterning than feedback about any one trial would be. Also, summary feedback necessarily involves strings of performance trials without feedback, which also tends to minimize corrections and stabilize performance relative to giving feedback after every trial. This less frequent feedback could reduce the learner's tendency to modify the movement unnecessarily, and thus generate more stable performance when feedback is eventually

withdrawn. This work provides an interesting link between the KR research and the newer thinking on kinematic feedback, but much needs to be done to resolve these various uncertainties about how summary feedback operates.

The lack of fading effects in this experiment tends to suggest that the KR and kinematic-feedback principles might be somewhat different, as the analogous effects in KR (Winstein & Schmidt, 1990; see Figures 2 and 4) were very powerful. Several explanations are possible. First, the tasks used are different, with the KR work being done with the spatial-temporal patterning task, and the kinematic feedback work being done with the coincident-timing task; thus, there are many variables that are not controlled between these two experiments. Even with the KR work, though, some of our recent studies suggest that the fading effect tends to be dominated by the effect of reduced relative frequency (Nicholson & Schmidt, 1991, Experiment 1), although there does tend to be a real, but small, advantage for fading schedules even when the relative frequency is held constant (Nicholson & Schmidt, 1991, Experiment 2). In addition, it is possible that the 5-trial Average kinematic feedback was nearly optimal for learning this task--just as the 5-trial summary KR was nearly optimal for learning in this task (Schmidt et al., 1990; see Figure 9)--perhaps leaving little opportunity for further improvement when this averaged kinematic feedback was then faded. Despite these doubts, there is certainly no evidence that the faded condition was worse for learning, and some suggestion that it was more effective, even though this condition had half the number of feedback reports as the Average condition received. This fading variable needs additional work before we can be certain about the role of fading in kinematic feedback paradigms.

Further Evidence Against the Specificity View

This kinematic feedback work provides additional evidence against the specificity (or similarity) hypothesis discussed earlier. Young's (1988) Experiment 1 contrasted several groups, two of which received KR after every trial in acquisition and retention. However, one group (Average) also received kinematic feedback about average backswing position after each 5-trial block (identical to the Average condition just discussed in conjunction with Figure 11), while the other condition received no added kinematic information (the KR-only group). After acquisition, all subjects were tested for learning on a retention test without kinematic information, where KR was always provided after each trial. Thus, the KR-only condition had identical feedback conditions in acquisition and retention, while the Average condition had kinematic feedback in acquisition but not in retention. The Average condition performed

considerably more effectively in retention than the KR-only condition, which was opposite to the expectations of a specificity view.

EMPIRICAL PRINCIPLES OF FEEDBACK

These results reinforce the desirability of a general learning-performance distinction (Salmoni et al., 1984) for the study of feedback, in that the performance levels seen at the end of an acquisition phase with feedback present are not necessarily indicative of performance capabilities of the subjects. Temporary effects associated with the presence of feedback enhance performance while feedback is present, but performance is altered markedly when these temporary factors are removed (on delayed no-KR retention tests), shown by often drastic reordering of the treatment conditions relative to acquisition. Clearly, the most effective performance conditions for learning are not necessarily those which produce the most effective performance in practice.

Decreasing Feedback's "Usefulness" Can Enhance Learning

A common theme in the work reported here is that feedback that is in one way or another made less "useful" to subjects for trial-to-trial adjustment seems to degrade performance during the acquisition phase, but to enhance learning as measured on several different kinds of retention tests. We have seen this with summary feedback, where more trials included in the summary reduces the learner's capabilities to use it to modify behavior. Also, reducing frequency of KR (and kinematic feedback) gives the learner less information during practice that can be used for modifying the next attempt(s). Average kinematic feedback does not give information for a string of trials, also providing less information for adjustment. Usually, this results in decrements in performance relative to every-trial feedback (an exception is average KP, which benefits performance). Yet the fact that all of these variables generally enhance performance in retention raise serious challenges to the usually accepted statements of the empirical principles of feedback (e.g., Adams, 1971; Bilodeau, 1966; Newell, 1977) and to the theories and conceptualizations of feedback that are dependent on these empirical principles as facts to be explained (e.g., Adams, 1971; Schmidt, 1975; Thorndike, 1927).

Traditional viewpoints about the functioning of feedback for learning have a difficult time with these findings. All of these existing viewpoints emphasize that, for learning to be maximized, the learner should be able to associate, link, or otherwise connect the production of

a given movement with its outcome in terms of meeting the environmental goal (expressed as KR). How these processes are conceptualized, and how feedback is thought to operate, differ markedly in the various accounts of learning, of course. For example, feedback has been said to produce learning by guiding the learner to the target behavior (e.g., Adams, 1971; Schmidt, 1975), by strengthening the bonds between the stimulus conditions and the goal response (Thorndike, 1927), or by increasing the capability to define input-output rules about limb control (e.g., schemas, see Schmidt, 1975). All of these views predict that making feedback more frequent, more precise, more immediate, and more informational will enhance learning. And, all of these views suggest that trials without feedback are either "neutral" for learning (e.g., Schmidt, 1975; Thorndike, 1927), and/or provide decrements in the capability for responding (e.g., Adams, 1971). Both of these general predictions are contradicted by the results presented here, in that conditions which make KR more "difficult" to use, or which withhold KR on a portion of trials, seem to increase--not decrease--retention capabilities.

Given this common feature in the various styles of explanation of how feedback functions for learning, the most important question coming from the work presented here is this: How can learners receiving less feedback, and/or feedback that is in various ways "difficult" to use informationally, learn more than subjects with clear error information after each trial? The reminder of this report is devoted to preliminaries to a theory of feedback that attempts to explain some of these empirical effects.

Specificity, or Similarity, Theories

One viewpoint that can account for at least some of our data here is the specificity view, which claims that the reduced "usefulness" of feedback in acquisition simply mimics the no-feedback retention tests often used. If so, there should be little surprise that acquisition conditions which match the test conditions should be most effective. However, in various places in this report, evidence against this view has been presented, and it is summarized briefly here.

In the relative frequency paradigms, we have found that, even if the retention test is conducted under conditions with feedback after each trial, reduced relative frequency (with fading) still produced more effective performance at retention than 100% feedback (Winstein, 1988; Winstein & Schmidt, 1990). Also, the relative frequency in acquisition does not interact with the relative frequency in the retention test, which would be expected by a specificity view. In the experiments on summary feedback, the acquisition groups with 10- and 15-trial

summaries should have performed better than the 5-trial summary groups on a no-feedback retention test, and yet the reverse effect was found (Schmidt et al., 1989, 1990). When subjects were transferred to one-trial summary conditions in retention, a 5-trial condition in acquisition was certainly no worse than, and tended to be better than, the 1-trial summary condition in acquisition; this was the case even with the retention test under 1-trial summary conditions (Schmidt et al., 1990). And, in the kinematic feedback paradigm, Young (1988; Young & Schmidt, 1991) found that the average condition (plus KR) was more effective for a KR-only retention test than a KR-only condition in acquisition, the latter having identical conditions in acquisition and retention. Finally, one clear prediction from a specificity view is that no feedback at all will be most effective for learning if learning is measured on a no-feedback retention test; this prediction is violated in several experiments here and elsewhere (e.g., Bilodeau et al., 1959; Trowbridge & Cason, 1932).

Also, the specificity hypothesis appears not to be supported in other paradigms. For example, blocked as opposed to random practice of a set of tasks is more effective for performance in acquisition, but is less effective for performance on delayed retention tests--regardless of whether the retention test is under blocked or random conditions [Lee & Carnahan, 1990; Shea & Morgan, 1979; see Lee (1988) or Magill & Hall (1990) for reviews]. Also, practicing a set of variations of a given task, as opposed to constant practice on one variation, is more effective for performance on a novel variant of this same task, regardless of whether the retention test is under variable or constant conditions (e.g., Shapiro & Schmidt, 1982). Similar findings are available in the verbal/cognitive domain as well (Bransford, Franks, Morris, & Stein, 1979; see Schmidt & Bjork, in press, for a review).

Overall, there may be some sense in which the similarity per se of the conditions in acquisition and retention is a factor in retention performance, but a specificity (or similarity) hypothesis holding that it is the primary factor cannot be taken seriously in view of the evidence presented here. Clearly, other ways to account for these effects need to be considered. In the next sections, new concepts underlying the use of feedback in practice and retention are outlined, and the evidence supporting them is given. These ideas form the foundations for a new theory of feedback use that has grown from the guidance hypothesis.

PRELIMINARIES TO A GUIDANCE THEORY FOR FEEDBACK

The guidance hypothesis, initial suggestions for which were given in preliminary form by Annett (1969), Holding (1965, 1968), and Lintern (1980), is a two-factor theory of feedback

use. First, a "positive" factor concerns the informational contribution that drives behavior toward its goal. Second, several "negative" decremental factors can be identified that cause the learner to become dependent on feedback in various ways, degrading performance at retention as we have seen.

Informational Properties of Feedback Direct Behavior

It has been long recognized that feedback provides an informational role for the learner that offers numerous benefits during practice when feedback is present. Most importantly, feedback informs about errors in goal achievement, and directs (or guides) the novice learner to make corrections on subsequent trials. In this way, feedback tends to hold the learner on target, signaling errors before they become too large, and triggering corrections when they are needed. Also, feedback serves a motivational (or "energizing") role, helping the learner to stay interested in the task, to work harder and bring more intellectual resources toward the solution of the movement problem, and to persist longer in the face of discouragement when skill levels are low (Salmoni et al., 1984). These factors are critically important for learning, and have been recognized as such for decades.

Most previous theorists have assumed in one way or another that feedback variations contributing to effective behavior were in some way interacting with these behaviors to produce permanent memories--either by strengthening bonds, by enhancing perceptual traces, or by building schemata. A critical expectation was that a movement and its feedback consequences should be very closely tied in time (S-R contiguity) for learning to be effective. However, the recent evidence presented here suggests that even radical dissociations between a given movement and its feedback--produced either by summarizing feedback, by delaying feedback in various ways from the behavior, or even by occasionally withholding feedback altogether--can be very effective for learning, leading to doubts that the literal pairing of feedback with behavior closely in time is a critical feature for learning after all.

Adams (1971) wrote that feedback operated primarily by driving the behavior toward the goal. Then, by quite separate processes, learning was accomplished by the strengthening of a perceptual trace through "laying down" a feedback-record of the sensory consequences of many nearly correct movements. Thus, for Adams, and for me (but for different reasons), this guiding property of feedback is not the most important process in learning. Rather, guidance merely serves to cause the subject to respond in such a way that the other critically important learning processes could occur. However, I reject Adams' notion that the strengthening of

perceptual traces, per se, is the basis for improved performance (see e.g., my arguments in Schmidt, 1975, 1982, 1988), primarily because of the strong evidence that much of movement behavior is open-loop in nature. But, even so, I support in general Adams' idea that feedback guides the learner during acquisition, thereby allowing other learning processes to operate that strengthen memory, as this idea fits the empirical data very well.

What are these learning processes? Certainly the most obvious of them is mere repetition, where feedback functions to drive the learner repeatedly to produce active behavior that approximates the movement's goal. Thus, it is the <u>production</u> (and/or planning) of this action, and its repetition over many subsequent trials, that serves to form a permanent memory. Feedback's role, in this way, is to drive the behavior toward the goal, and to ensure that the behavior continues to be effective, signalling occasional needs for corrections when the learner drifts off target. But feedback is not needed after every trial so long as the movement is reasonably close to goal. In fact (as discussed below), feedback on every trial may even be detrimental to establishing memory because it disrupts the repetition of the movement from trial to trial.

Inherent in the idea that repetition is critical for learning are two possibilities. It may be the actual <u>production</u>, or <u>execution</u>, of the movement that is essential for learning, with the intrinsic feedback from proprioception, vision, audition, etc., being critical for the gains due to repetition. But this notion seems contradicted by several lines of evidence. First, the literature on the remarkable gains in learning via mental practice (e.g., Heuer, 1986), where no discernible movement behavior can be observed, raises doubts that movement production is critical. In view of the earlier, traditional viewpoints for feedback that demanded the linkage of a movement and its feedback, this literature on mental practice was difficult to understand, in that no movement--and hence no feedback--was even produced, and thus no learning should occur. Second, disrupting trial-to-trial repetitions has been shown to be beneficial in the learning of both verbal skills (e.g., Landauer & Bjork, 1979; Cuddy & Jacoby, 1982) and motor actions (Lee, 1988; Lee & Magill, 1983; Magill & Hall, 1990).

Thus it is more likely that the critical factor in learning could be the planning of the action, including the "retrieval practice" (Bjork, 1979) associated with accessing materials in memory, and the movement programming and parameterization that readies the movement apparatus for movement. These preparatory processes have been studied a great deal, but their role in movement learning has not been considered until recently, and thus their importance must remain speculative here. But the suspicion raised by the evidence about movement feedback is that preparation is far more important for learning than has been realized. In fact, it is possible

that frequent feedback has its decremental roles on learning because it make the next movement in the sequence "too easy" to plan, more or less as blocked practice does. If so, this might be a way to understand the counterintuitive effects of infrequent feedback and randomized practice--which facilitate learning relative to frequent feedback and blocked practice, respectively. Unfortunately, these various ideas cannot be clearly separated on the basis of the evidence presently available, although the potential for movement preparation as the critical variable seems somewhat stronger.

Guidance Produces a Reliance on Feedback in Acquisition

Our initial ideas with respect to the guidance hypothesis were that frequent, very informative feedback in acquisition tends to encourage the subject to develop an over-reliance on it, which can be manifested in several ways as discussed later here. This dependence on feedback is acceptable so long as it is present in acquisition, because performance is effective due to feedback's powerful trial-to-trial guiding effects. However, when feedback is withdrawn in retention, this over-reliance is revealed as a marked disruption in performance as compared to conditions with less "useful" feedback in acquisition. This dependency on feedback can be conceptualized in several ways, discussed next.

Feedback Becomes a "Part" of the Task

When feedback is presented after each trial, the subject learns to operate as if the extrinsic feedback is a fundamental "part" of the task, and comes to expect it as informational support for the action on subsequent trials. In this sense, strictly extrinsic feedback tends to become "intrinsic" feedback for the learner, as if it has become "inherent" or "natural" to the production of that action. If so, when this critical, learned, informational support is removed or altered in a retention test, performance suffers. This is just what seems to happen in the literature on physical guidance (e.g., Holding, 1965, 1968), where the guidance serves as a kind of "crutch" to support behavior--useful so long as it is present, but detrimental to movement functioning (even compared to no-guidance control conditions) when it was removed at test.

The notion that feedback becomes a "part" of the task is supported in experiments by Proteau and Marteniuk (1987). Subjects learned an aiming movement with vision present or not, and then both groups transferred to a retention test without vision; even though vision in acquisition greatly facilitated performance there, on the no-vision retention test the group that

learned without vision was more effective. Apparently, subjects with vision in acquisition acquired a vision-dependency--vision becoming a "part" of the task--and performance deteriorated when it was removed in retention (see also Griffith, 1931). Admittedly, this visual feedback provides information during the action as well as after it (unlike KR does). But these data provide encouragement to the idea that feedback could be learned as a "part" of the task also, with decrements in performance when it is withdrawn.

Enhanced Learning of Error-Detection Capabilities

Frequent feedback during acquisition might not always act to become a "part" of the task, but rather might act to interfere with, or even to block, other important information processing activities that occur without feedback. Frequent feedback is useful for maintaining effective performance, so the learner is presumably not motivated to engage in extra, mentally effortful information-processing activities during acquisition. One of these processes concerns spontaneous efforts by subjects to detect their own errors after each movement. In one of our earlier experiments (Schmidt & Shapiro, 1986), we interviewed subjects after they had practiced relatively simple ballistic-timing tasks for an hour. A majority of subjects said that they spontaneously attempted to determine their own errors after most trials in acquisition, presumably by noting how the movement sounded and felt. This attention to intrinsic feedback seems necessary for the subjects to develop a reference of correctness based on the past experience (Schmidt, 1975; Schmidt & White, 1972). Then, subjects with this capability can detect their own errors later by comparing the response-produced feedback on a given trial with their reference of correctness, providing subjective information about errors that can effectively substitute for the withheld KR.

The difficulty with frequent KR in acquisition is that it tends to block the need for the subject to process movement information in this way. Of course, while waiting for extrinsic feedback, the learner is free to estimate his or her own errors via this comparison of response-produced feedback and the reference of correctness. But why should learners engage in these relatively effortful analyses when the experimenter always provides the answer "for free" in a few seconds? The notion, then, is that frequent feedback, because it gives so much information, effectively blocks subjective efforts to process movement-produced feedback, resulting in retarded development of error-detection capabilities. Performance suffers when feedback is withdrawn at test, as any errors that are produced are not recognized and corrected, leading to continued errorful movements.

Evidence from objective-subjective correlations. We have already discussed evidence from the summary-KR paradigm which favors this view. Here, the 5-trial summary group that was most effective in retention (compared to 1-, 10, and 15-trial groups) also had the strongest capability to detect their own errors, as measured by within-subject correlations between actual and estimated scores. However, these effects were relatively small, error-detection performance was in any case not very accurate, and these data are vulnerable to the argument that the objective-subjective correlations are themselves correlated with overall scores for the task. So for several reasons, these data are not particularly compelling.

The instantaneous-KR paradigm. However, additional evidence from a different paradigm provides converging evidence for the notion that frequent KR may block error-detection processes. Swinnen, Schmidt, Nicholson, and Shapiro (1990) assumed that, if KR is presented instantaneously after a movement, then the learner's motivation to process response-produced proprioceptive and auditory movement information in order to evaluate their own errors would be essentially eliminated. This is so because the result of that processing (i.e., the subjective estimate of the score on that trial) would already have been provided by the experimenter as KR. This notion leads to the interesting prediction that giving instantaneous KR will be less effective for developing error-detection processes than will KR that is delayed somewhat. And, instantaneous KR should be detrimental for performance on retention tests without feedback. This is, of course, counter to most theoretical expectations about immediate feedback and learning that have been present for several decades (Bilodeau, 1966, 1969; Tarpy, 1975).

There were three similar experiments (see also Swinnen, 1988), all giving essentially the same result, but the third one using the more "complex" coincident-timing apparatus discussed earlier is presented here. One group of subjects received KR about the score on the task 320 ms after the coincidence-point was reached, which gave the subjective impression of feedback being nearly instantaneous. Another group received the same information delayed by 3 s--presumably enough time to allow subjective error-detection processes to occur. Subjects received two days of acquisition (90 trials per day) under these differential KR-delay conditions, and then received no-KR retention tests after 10 min, two days, and four months. The average scores are shown in Figure 12. There were essentially no differences between groups on Day 1. But by Day 2, the delayed KR group was generating approximately 25% larger scores than the instantaneous group. This difference persisted into the various retention tests, and was present even after a four-month retention interval. Clearly, the instantaneous KR degraded learning in this task. Our interpretation was that it did so by blocking error-detection

processes necessary for maintenance of performance on the no-KR tests.

Figure 12 about here

Preventing Maladaptive Short-Term Corrections

One of the most provocative notions contributing to the present ideas stemmed from Lavery's (1962) findings in the summary-KR paradigm. He showed that a group with summary KR, but which also had KR after each trial (the Both group in Figure 7) showed performance and learning essentially identical to that of the Immediate KR group. To us, this suggested that it was not so much the benefits of the summary-KR reports per se that were influencing learning, but rather the detrimental effects caused by the KR after each trial. The suggestion that frequent KR could be harmful for learning seemed to threaten the very core of traditional belief about feedback and learning.

Lavery's work, plus several other findings mentioned previously, suggest that many of the benefits of reduced "usefulness" of feedback stem from the fact that the subject is prevented, or at least not encouraged, to make corrections in the movement on every trial. On the other hand, subjects with feedback after each trial are encouraged to make "maladaptive short-term corrections"[5] which might make the movements somewhat more accurate in the short term (but less stable, as shown by Sherwood, 1983). But this frequent feedback could interfere with the long-term learning of these actions. If the behavior is changed after each movement as a result of frequent feedback, the subject does not have the opportunity to repeat a given action, and these short-term corrections act to prevent the establishment of smooth, stable patterns. This is particularly important if a critical factor in learning is repetition, as was suggested earlier. Conversely, learners who must produce long strings of trials without feedback tend to make the same, less generally correct, movement on every trial, which perhaps results in more movement stability in practice. Contrary to the earlier views that errors should be corrected after every attempt, the present view is that such short-term modifications are actually maladaptive when the long-term goal of effective retention is considered.

Perhaps surprisingly, these notions agree in a general way with common practices used by many excellent teachers and coaches. Particularly in the early stages of practice, coaches often allow the learner to produce a sequence of actions without feedback. This allows the learner to produce the same behavior repeatedly, which "filters out" the trial-to-trial variations, so that the typical movement pattern can be perceived by the instructor as a basis for later giving feedback. Also, when the feedback is finally given, the learner can relate it to stable characteristic of his or

her movement patterning, rather than to variations that may have been spurious and would not have been repeated again in any case. Also, many coaches feel that it is more important to achieve the proper form (or pattern of action) first, and then later to provide feedback about goal achievement that will focus the learner on the outcome. According to this view, in shooting a basketball, a player should first achieve the proper leg and arm coordination, wrist extension at ball release, and the proper arc on the ball's trajectory. Only when this pattern learning is achieved should the focus be turned to whether or not the basket was made. In this sense, feedback about goal achievement might force maladaptive short-term corrections in the pattern; this might result in more baskets being made in practice, but would actually disrupt learning of the essential patterns that will be critical for retention.

Additional Benefits of Withheld Feedback

Several other potential benefits of withholding feedback or reducing its "usefulness" can be suggested, although there is less empirical evidence for these notions than for the earlier ones. First, long strings of no-feedback trials may allow the learner to drift away from the target behavior sufficiently that, when feedback is finally presented, the deviations of the movement from its goal are readily apparent to the learner, and the nature of the correction needed will be more obvious. Without any no-feedback trials, behavior will be so heavily guided that it will not deviate from the target, making errors in the movement less easily detected when feedback is given (Winstein, 1988). Along similar lines, larger errors in performance generated from strings of no-feedback trials will produce feedback reports that are themselves more meaningful, as the errors will be more easily described if they are large. Finally, if variations around a target behavior are beneficial for learning, as the literature on schema development suggests (Shapiro & Schmidt, 1982; Schmidt, 1975), then withholding feedback might potentiate at least one form of this variability--bias away from the target behavior. Other hypotheses are possible as well.

Some Points of Contact With Reinforcement Research

There are several apparent similarities between the present ideas about feedback and the large body of work on reinforcement schedules in animals (and sometimes humans). These paradigms for animal reinforcement studies and human motor learning studies are vastly different, of course. The reinforcement work has always been concerned with the elicitation of

previously learned behavior (bar pressing, pecking, etc.) when a given stimulus is presented, and the animal's problem is the choice among several learned actions when the stimulus is presented. By contrast, the human motor learning work is concerned with the acquisition of new movement behaviors, where there is usually little question of which movement to make, but rather how to make a given movement faster, more accurately, and so on. I know of no animal work where novel motor behaviors are acquired with practice, although such capabilities must surely exist (e.g., circus dogs learn to do somersaults). Even so, in some ways, the findings we have generated for feedback seem parallel with those in the animal work, suggesting that the principles of feedback for humans might be the same as for reinforcement in animals. Some of these similarities and differences are mentioned next.

Partial Reinforcement, Faded Reinforcement, and Extinction

One example is the relative frequency paradigm mentioned earlier, which is similar to the partial-reinforcement and faded-reinforcement paradigms in conditioning work. With partial reinforcement, reducing the proportion of reinforcements generally degrades performance in training when reinforcement is being presented, but produces greater resistance to extinction (i.e., no-reinforcement) performance later on. In the work on KR relative frequency, reduced relative frequency degrades performance in acquisition, but enhances no-KR retention performance, the later being analogous to extinction trials in the animal work. Also, faded reinforcement, where the probability of reinforcement is gradually reduced across training, generally enhances performance in extinction trials. This is strongly analogous to our faded KR paradigm, where fading enhances no-KR retention performance (Figure 2). While these strong similarities suggest common mechanisms and theoretical notions, a closer examination reveals differences between these paradigms as well.

One major hypothesis in the reinforcement work is the specificity (or similarity) view mentioned earlier. Partial, faded reinforcement is said to be effective because it tends to simulate the extinction conditions (Tarpy, 1975). Thus our findings here, which argue against a specificity view for KR and learning, also tends to be inconsistent with this interpretation of the reinforcement work. For example, the effects of reduced KR relative frequency were not specific to the KR conditions on the retention test (Figure 3), but were as large for 100%-KR as they were for no-KR retention tests. The analogous effect in the reinforcement work would probably not occur, as the 100% reinforcement conditions at retention would probably overshadow the effect of training conditions. In addition, the evidence on faded feedback

shows that this variable is more effective than every-trial feedback even for 100% KR retention tests (Figure 4), a result which would seem unlikely to occur in the animal work (Tarpy, 1975). Thus, outcomes of experiments in which the feedback conditions are varied in acquisition, but with 100% KR retention tests, do not parallel those from the reinforcement work very closely.

Instantaneous KR and Immediate Reinforcement

Another apparent similarity in the animal reinforcement and human feedback work concerns the studies on delay of feedback or reward. A strong finding from the reinforcement work is that delaying the reward in time from goal behavior degrades both performance in training and on retention tests (Tarpy, 1975). In fact, if this delay is long enough, the animal will never reliably produce the goal behavior. This analogous effect has been studied at length in the human KR literature, with experimenters expecting to find that delayed KR would degrade learning of movement behaviors. These effects never materialized for performance during the acquisition phase, with many experiments showing no effect of delayed feedback at all. And, no effects on learning could be found when various retention tests without feedback were used (see Salmoni et al., 1984, for a review). This difference in findings has been taken by many as evidence that the human motor literature and the animal reinforcement literature were not based on common principles (e.g., Adams, 1968, 1971).

An additional difference between these two lines of work is evident here the instantaneous feedback paradigm here (Figure 12). In the animal work, giving reinforcement very quickly generally aids performance and learning, whereas instantaneous feedback degrades motor learning. It is possible that the addition of extra information-processing activities for human learning, such as the attention to response-produced feedback and the learning of error-detection mechanisms, are not present (or at least to the same extent) in animal subjects, leading to the differences in findings. Overall, while our findings with several variations of feedback scheduling appear to be fundamentally similar to analogous findings in animal reinforcement work, there are too many instances where these findings do not operate in the same ways for us to take this parallel very seriously.

PRACTICAL IMPLICATIONS FOR TRAINING

The findings from this project have numerous implications for practical training settings. The variations in feedback which produce the most effective learning can often be produced

with only minor variations in training methods already in place. In present training settings, summary KR could as easily be provided in riflery practice, and it might even be easier to provide than immediate KR. Certainly KR could be withheld on certain trials to create reduced relative frequency, with this relative frequency being high in early practice when the learner is just acquiring the skills, and reduced later as proficiency increases. And, feedback does not have to be delivered instantaneously, again simplifying the problem for the design of training settings. These feedback variations, with some modification, could be employed in land-navigation training settings as well. In addition, the benefits in terms of long-term retention appear to be relatively large, suggesting that these changes will be relatively cost-effective ways to enhance Army training.

A central point is that variations in feedback that enhance performance in acquisition are not necessarily the same ones that will produce optimal performance in delayed retention tests (see Schmidt, 1991). This is a particularly important problem in Army settings such as marksmanship training. This is a skill for which the criterion is not necessarily success in a training session, but rather is success in a delayed retention test (e.g., on a battlefield). Further, these conditions often involve degraded feedback, often with poor lighting or dusty, smoky conditions, so that the soldier cannot determine very easily whether or not a target has been hit. The laboratory experiments described above, where success in training can be evaluated in terms of performance on a delayed no-KR retention test, share many of the same features of this kind of Army training. Further, they suggest the possibility that efforts to provide high-fidelity feedback in Army training (and in simulators) could be overdone; this might allow the learner to be very proficient <u>in training</u> because of various temporary phenomena, but might render him or her relatively ineffective on an important <u>criterion</u> test to be performed later.

This point has strong implications for the ways in which trainers are trained as well. An industrious and well-intentioned trainer will naturally attempt to do everything possible in practice to facilitate performance, as this is what the job is assumed to entail. But many of the techniques that would be used--such as frequent feedback, instantaneous feedback, blocked practice, and the like--which have strong performance-facilitating influences during practice do not produce effective long-term retention. Thus, the very methods used by the most dedicated trainer could be detrimental to important Army training goals, the most important of which is long-term retention. This suggests the need to educate those responsible for training with respect to these other important goals of practice, and how they can be best attained with effective training methods.

These findings also have important implications for simulator design, such as used in

riflery and vehicle control. It has usually been assumed, based presumably on earlier viewpoints about feedback and learning, that effective simulators must have feedback which is of high "fidelity" and high frequency. This is understandable, as feedback has been designed with the primary criterion of enhanced performance in the simulator. Our data raise the likelihood that this feedback can be overdone, particularly when the artificial feedback from the simulator is removed in the actual criterion setting. It is perhaps unwise to generalize our results too far in this direction, but these data warrant some skepticism about the general beliefs about the role of feedback quality in simulators.

Finally, these variations in feedback have their effects evident mainly in retention tests. Of course, variations in training which increase readiness and decrease the need for retraining and refresher courses should provide considerable savings in time and resources. Also, the variables under study here appear to have their effects in conditions for which the feedback is withdrawn or at least degraded. In riflery, for example, shooting is often done at night, or under smoky or dusty conditions, making outcome information difficult. This stresses the need for training which fosters the subject's internal capabilities for responding, and which do not require the support from feedback at the time that the skill will eventually be used.

SUMMARY

The evidence presented here has revealed in several ways the inadequacy of present principles of feedback for learning, and of course the theories that have been generated to account for them. These views have always stressed that information that was more immediate, more precise, more frequent, and that provided more information about the errors in a movement, would be most effective for learning. This viewpoint might be essentially correct as far as performance during the practice phase is concerned (but there are exceptions to this as well--e.g., Figures 11 and 12), but it fails to consider the distinction between the temporary performance effects of feedback and the longer-term learning effects. When these latter effects are revealed in various kinds of retention and transfer tests, a number of findings emerge to suggest that these older ideas about feedback need to be revised. Certainly, the most provocative one of them is the persistent finding here that providing feedback less often, and/or in ways that make it "difficult" for the learner to relate a given movement and its feedback, seem to degrade performance in practice but to enhance learning as measured on a variety of long-term retention tests. Thus, the problem becomes one of how this general set of findings could be explained.

The preliminaries to a guidance theory of feedback outlined here represents one way to account for this effect, proposing "positive" and "negative" effects of feedback that operate together during acquisition. It is well known that feedback provides guidance to learners in practice, and drives the behavior toward the task goal by systematically reducing errors. But at the same time, the very utility of this guidance creates a kind of reliance on feedback, which I conceptualize as allowing the feedback to become a "part" of the task, as blocking the subjective processing of intrinsic information that could allow the learner to develop more effective and stable memory representations (e.g., error-detection mechanisms), or as creating too many trial-to-trial maladaptive short-term corrections that block the acquisition of response stability. Evidence is provided for all of these processes, but it is far too early to determine if they will be continue to be supported in future experiments. This set of processes is not exhaustive, as the existence of other such processes are possible, even probable.

Finally, these empirical results, and this first attempt to account for them, provide a new framework for feedback and learning that seems useful in many practical training settings. The realization that feedback has several effects at once would seem to emphasize the need for more thought and planning to be directed to feedback administration--focusing on what information is provided, and how it is scheduled. More generally, these additional complexities for feedback combine with new findings in the organization of practice (e.g., Lee, 1988; Magill & Hall, 1990; Schmidt, in press; Schmidt & Bjork, in press) to encourage a renewed research emphasis in understanding the surprisingly diverse collection of processes operating in what we commonly call "practice."

REFERENCES

Adams, J.A. (1968). Response feedback and learning. Psychological Bulletin, 70, 486-504.

Adams, J.A. (1971). A closed-loop theory of motor learning. Journal of Motor Behavior, 3, 111-150.

Adams, J.A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin, 101, 41-74.

Annett, J. (1969). Feedback and human behavior. Middlesex, England: Penguin.

Armstrong, T.R. (1970). Training for the production of memorized movement patterns. Technical Report No. 26, Human Performance Center, University of Michigan.

Arps, G.F. (1920). Work with KR versus work without KR. Psychological Monographs, 28 (No. 125).

Bilodeau, E.A., & Bilodeau, I.M. (1958). Variable frequency knowledge of results and the learning of simple skill. Journal of Experimental Psychology, 55, 379-383.

Bilodeau, E.A., & Bilodeau, I.M., & Schumsky, D.A. (1959). Some effects of introducing and withdrawing knowledge of results early and late in practice. Journal of Experimental Psychology, 58, 142-144.

Bilodeau, I.M. (1956). Accuracy of a simple positioning response with variation in the number of trials by which KR is delayed. American Journal of Psychology, 69, 434-437.

Bilodeau, I.M. (1966). Information feedback. In E.A. Bilodeau (Ed.), Acquisition of skill (pp. 255-296). New York: Academic Press.

Bilodeau, I.M. (1969). Information feedback. In E.A. Bilodeau, (Ed.), Principles of skill acquisition (pp. 255-285). New York: Academic Press.

Bjork, R.A. (1979). Retrieval practice. Paper presented at the Conference on Developmental and Experimental Approaches to Memory, Ann Arbor, MI.

Bransford, J.D., Franks, J.J., Morris, C.D., & Stein, B.S. (1979). Some general constraints on learning and memory research. In L.S. Cermak & F.I.M. Craik (Eds.), Levels of processing in human memory (pp. 331-354). Hillsdale, NJ: Erlbaum.

Crawley, S.L. (1926). An experimental investigation of recovery from work. Archives of Psychology, 13, 26.

Cuddy, L.J., & Jacoby, L.L. (1982). When forgetting helps memory. Analysis of repetition effects. Journal of Verbal Learning and Verbal Behavior, 21, 451-467.

Elwell, J.L., & Grindley, G.S. (1938). The effect of knowledge of results on learning and performance. British Journal of Psychology, 29, 39-53.

Fowler, C.A. & Turvey, M.T. (1978). Skill acquisition: An event approach for the optimum of several variables. In G.E. Stelmach (Ed.), Information processing in motor control and learning (pp. 1-40). New York: Academic Press.

Gentile, A.M. (1972). A working model of skill acquisition with application to teaching. Quest, 17, 3-23.

Gentner, D.R. (1987). Timing of skilled motor performance: Tests of the proportional duration model. Psychological Review, 94, 255-276.

Gopher, D., Williges, B.H., & Williges, R.C. (1975). Varying the type and number of adaptive variables in continuous tracking. Journal of Motor Behavior, 7, 159-170.

Griffith, C.R. (1931). An experiment on learning to drive a golf ball. The Athletic Journal, 11, 11-13.

Guthrie, E.R. (1952). The psychology of learning. New York: Harper and Row.

Henry, F.M. (1968). Specificity versus generality in learning motor skill. In R.C. Brown and G.S. Kenyon (Eds.), Classical studies on physical activity (pp. 331-340). Englewood Cliffs, NJ: Prentice-Hall.

Heuer, H. (1986). Wie wirkt mentale Uebung? [How does mental practice operate?] Psychologische Rundschau, 36, 191-200.

Heuer, H., & Schmidt, R.A. (1988). Transfer of learning among motor patterns with different relative timing. Journal of Experimental Psychology: Human Perception and Performance, 14, 241-252.

Ho, L., & Shea, J.B. (1978). Effects of relative frequency of knowledge of results on retention of a motor skill. Perceptual and Motor Skills, 46, 859-866.

Holding, D.H. (1965). Principles of training. Oxford: Pergamon Press.

Holding, D.H. (1968). Training for skill. In D. Pym (Ed.), Industrial society: Social sciences in management (pp. 134-151). Baltimore: Penguin.

Hull, C.L. (1943). Principles of behavior. New York: Appleton-Century-Crofts.

Johnson, R.W., Wicks, G.G., & Ben-Sira, D. (1981). Practice in the absence of knowledge of results: motor skill retention. Unpublished manuscript, University of Minnesota. (Cited in Schmidt, 1982).

Landauer, T.K., & Bjork, R.A. (1979). Optimum rehearsal patterns and name learning. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), Practical aspects of memory (pp. 625-632). London: Academic Press.

Lavery, J.J. (1962). Retention of simple motor skills as a function of type of knowledge of results. Canadian Journal of Psychology, 16, 300-311.

Lavery, J.J. (1964). Retention of a skill following training with and without instructions to retain. Perceptual and Motor Skills, 18, 275-281.

Lavery, J.J., & Suddon, F.H. (1962). Retention of simple motor skills as a function of the number of trials by which KR is delayed. Perceptual and Motor Skills, 15, 231-237.

Lee, T.D. (1988). Transfer-appropriate processing: A framework for conceptualizing practice effects in motor learning. In O. Meijer & K. Roth (Eds.), Complex movement behavior: "The" motor-action controversy (pp. 210-215). Amsterdam: North-Holland.

Lee, T.D., & Carnahan, H. (1990). When to provide knowledge of results during motor learning: Scheduling effects. Human Performance, 3, 87-105.

Lee, T.D., & Magill, R.A. (1983). The locus of contextual interference in motor skill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 730-746.

Lee, T.D., White, M.A., & Carnahan, H. (1990). On the role of knowledge of results in motor learning: Exploring the guidance hypothesis. Journal of Motor Behavior, 22, 191-208.

Lintern, G. (1980). Transfer of landing skill after training with supplementary visual cues. Human Factors, 22, 81-88.

Magill, R.A., & Hall, K.G. (1990). A review of the contextual interference effect in motor skill acquisition. Human Movement Science, 9, 241-289..

McGuigan, F.J. (1959). The effect of precision, delay, and schedule of knowledge of results on performance. Journal of Experimental Psychology, 58, 79-80.

Newell, K.M. (1974). Knowledge of results and motor learning. Journal of Motor Behavior, 6, 235-244.

Newell, K.M. (1977). Knowledge of results and motor learning. Exercise and Sport Sciences Reviews, 4, 195-228.

Newell, K.M., & McGinnis, P.M. (1985). Kinematic information feedback for skilled performance. Human Learning, 4, 39-56.

Nicholson, D.E., & Schmidt, R.A. (1991). Feedback schedules to enhance training effectiveness. Proceedings of the Human Factors Society, 35th Annual Meeting. Santa Monica, CA: Human Factors Society.

Proteau, L., & Marteniuk, R.G. (1987). On the ability of visual information about the ongoing limb and near-body environment in the learning and control of an aiming movement. Paper presented at the Canadian Society for Psychomotor Learning and Sport Psychology, Banff, 1987.

Salmoni, A.W., Schmidt, R.A., & Walter, C.B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. Psychological Bulletin, 95, 355-386.

Schmidt, R.A. (1975). A schema theory of discrete motor skill learning. Psychological Review, 82, 225-260.

Schmidt, R.A. (1982). Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics Publishers.

Schmidt, R.A. (1985). The search for invariance in skilled movement behavior. Research Quarterly for Exercise and Sport, 56, 188-200.

Schmidt, R.A. (1988). Motor control and learning: A behavioral emphasis (2nd ed.). Champaign, IL: Human Kinetics Publishers.

Schmidt, R.A. (1991). Motor learning and performance: From principles to practice. Champaign, IL: Human Kinetics Publishers.

Schmidt, R.A. (in press). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G.E. Stelmach & J. Requin (Eds.), Tutorials in motor neuroscience. Dordrecht: Kluwer.

Schmidt, R.A., & Bjork, R.A. (in press). New conceptualizations of practice. Common principles in three paradigms suggest new concepts for training. Psychological Science.

Schmidt, R.A., Lange, C.A., & Young, D.E. (1990). Optimizing summary knowledge of results for skill learning. Human Movement Science, 9, 325-348.

Schmidt, R.A., & Shapiro, D.C. (1986). Optimizing feedback utilization in motor skill training. Technical Report No 1/86, Motor Control Laboratory, UCLA. (Contract No. MDA903-85-K-0225, U.S. Army Research Institute, Basic Research).

Schmidt, R.A., Shapiro, D.C., Winstein, C.J., Young, D.E., & Swinnen, S. (1987). Feedback and motor skill training: Relative frequency of KR and summary KR. Technical Report No. 1/87, Motor Control Laboratory, UCLA. (Contract No. MDA903-85-K-0225, U.S. Army Research Institute, Basic Research). AD# A191559

Schmidt, R.A., & White, J.L. (1972). Evidence for an error detection mechanism in motor skills: A test of Adams' closed-loop theory. Journal of Motor Behavior, 4, 143-153.

Schmidt, R.A., & Young, D.E. (1991). Methodology for motor learning: A paradigm for kinematic feedback. Journal of Motor Behavior, 23, 13-24.

Schmidt, R.A. Young, D.E., Swinnen, S., & Shapiro, D.C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 352-359.

Shapiro, D.C., & Schmidt, R.A. (1982). The schema theory: Recent evidence and

developmental implications. In J.A.S. Kelso & J.E. Clark (Eds.), The development of motor control and coordination (pp. 113-150). New York: Wiley.

Shea, J.B., & Morgan, R.L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. Journal of Experimental Psychology: Human Learning and Memory, 5, 179-187.

Sherwood, D.E. (1983). The impulse-variability model: Tests of major assumptions and predictions. Unpublished doctoral dissertation, University of Southern California.

Sherwood, D.E. (1988). Effect of bandwidth knowledge of results on movement consistency. Perceptual and Motor Skills, 66, 535-542.

Swinnen, S. (1988). Knowledge of results delay activities and motor learning. Unpublished doctoral dissertation, Catholic University of Leuven, Belgium

Swinnen, S., Schmidt, R.A., Nicholson, D.E., & Shapiro, D.C. (1990). Information feedback for skill acquisition: Instantaneous knowledge of results degrades learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 706-716.

Tarpy, R.S. (1975). Basic principles of learning. Glenview, IL: Scott, Foresman.

Thorndike, E.L. (1927). The law of effect. American Journal of Psychology, 39, 212-222.

Tolman, E.C. (1932). Purposive behavior in animals and men. New York: Century.

Trowbridge, M.H., & Cason, H. (1932). An experimental study of Thorndike's theory of learning. Journal of General Psychology, 7, 245-260.

Tulving, E., & Thomson, D.M. (1971). Retrieval processes in recognition memory: Effects of associative context. Journal of Experimental Psychology, 87, 116-124.

Williges, B.H., & Williges, R.C. (1977). Learner-centered versus automatic adaptive motor skill training. Journal of Motor Behavior, 9, 325-331.

Winstein, C.J. (1988). Relative frequency of information feedback in motor performance and learning. Unpublished doctoral dissertation, UCLA.

Winstein, C.J., & Schmidt, R.A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 677-691.

Wulf, G., & Schmidt, R.A. (1988). Variability in practice: Facilitation in retention and transfer through schema formation or context effects? Journal of Motor Behavior, 20, 133-149.

Wulf, G., & Schmidt, R.A. (1989). The learning of generalized motor programs: Reducing the relative frequency of knowledge of results enhances memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 352-359.

Young, D.E. (1988). Knowledge of performance and motor learning. Unpublished doctoral dissertation, University of California, Los Angeles.

Young, D.E., & Schmidt, R.A. (1990). Units of motor behavior: Modifications with practice and feedback. In M. Jeannerod (Ed.), Attention and performance XIII (pp. 763-795). Hillsdale, NJ: Erlbaum.

Young, D.E., & Schmidt, R.A. (1991). The role of augmented kinematic feedback in motor learning. Manuscript in preparation, UCLA.

ACKNOWLEDGEMENTS

Thanks to all of the many students and employees who have helped in this work. They are Dawn Asano, Michael Brennan, Carolee Winstein, Lorie Gunderson, Claudia Lange, Diane Nicholson, Annie Pauzie, Jose Pizano, Bob Rosen, Mike Shires, Stephan Swinnen, Normand Teasdale, Chuck Walter, Andy Wong, Douglas Young, Ray Young, and the hundreds of students at UCLA who served as subjects in our experiments. Particular appreciation is extended to my colleague in the earlier stages of this work, Diane Shapiro. Thanks also to Timothy D. Lee, Richard A. Magill, and Tony Sparrow for helpful comments on an earlier draft.

FOOTNOTES

1. In this report, I use the term "feedback" to mean all of these forms of extrinsic, augmented, post-response information.

2. Actually, this procedure was first used in an attempt to minimize trial-to-trial variations in so that EMG patterning would be more stable and clearly seen (Sherwood, 1983).

3. These group labels were used by Schmidt (1982, 1988), and do not appear in Lavery's report.

4. We make no claim here to provide an acceptable operational definition of the notion of "complexity." However, as compared to the ballistic-timing task, this task had more movement patterns that could be produced, involved a moving display, required a weighted combination of spatial accuracy and velocity, etc. Perhaps most importantly, the task performances do not asymptote for 900 trials or more (Young & Schmidt, 1990) in comparison to the rather rapid improvements seen in Figure 8, for example. By these criteria, at least, the coincident-timing task seemed to be more "complex" than the ballistic-timing task.

5. Thanks to Robert A. Bjork, who suggested this term in discussions about this work.

FIGURE CAPTIONS

Figure 1. Goal position-time pattern (trace ending at 800 ms) and a subject's attempt to produce it (extended trace) superimposed as feedback (from Winstein, 1988; Winstein & Schmidt, 1990).

Figure 2. Mean RMS error in spatial-temporal patterning for two days of acquisition (12 trials/block) under 100% KR or 50%-faded KR, and on a no-KR retention test on the third day (from Winstein, 1988; Winstein & Schmidt, 1990).

Figure 3. Mean RMS errors in spatial-temporal patterning on a 10-min retention test as a function of the relative frequency of KR in acquisition (100% or 33%) and retention (0%, 33%, 66%, or 100%) phases (from Winstein, 1988; Winstein & Schmidt, 1990).

Figure 4. Mean RMS error in spatial-temporal patterning for two days of acquisition (12 trials/block) under 100% KR or 50%-faded KR, and on a one-day retention test with 100% KR (from Winstein, 1988; Winstein & Schmidt, 1990).

Figure 5. Mean absolute constant error in relative timing on three variations (averaged together) of a timing task during acquisition under either 100% or 67% relative frequencies, and on a novel variation of this task under no-KR conditions after 10 min or 2 days (from Wulf & Schmidt, 1990).

Figure 6. Mean absolute error (6 trials/block) in relative timing on three versions (a, b, c) of a timing task during acquisition, and for the b-version on a no-KR retention test after 10 min or 2 days; for the b-version in acquisition, feedback was never given for the no-KR group and was always given for the KR group (from Wulf & Schmidt, 1990).

Figure 7. Mean percent correct (20 trials/day) in three tasks (averaged together) for Summary KR, Immediate KR, and Both conditions in acquisition, and on six delayed no-KR retention tests (from Lavery, 1962).

Figure 8. Mean absolute constant errors in ballistic timing (15 trials/block) as a function of the summary-KR length in acquisition, and on 10-min and 2-day no-KR retention tests

(from Schmidt, Young, Swinnen, & Shapiro, 1989).

Figure 9. Mean score (arbitrary units, 15 trials/block) on a coincident-timing task as a function of the summary-KR length in acquisition, and for 10-min and 2-day no-KR retention tests (from Schmidt, Lange, & Young, 1990).

Figure 10. Mean score (arbitrary units) on a coincident-timing task as a function of the summary-KR length (1 versus 5 trials) in acquisition (10 trials/block), and on retention tests (6 trials/block) with KR after each trial (from Schmidt, Lange, & Young, 1988).

Figure 11. Mean score (arbitrary units) on a coincident-timing task in acquisition (20 trials/block) as a function of variations in kinematic feedback (KP), and on no-KP retention tests after 1 day and 1 week; KR about the score was always present after each trial in acquisition and retention (from Young, 1988; Young & Schmidt, 1990).

Figure 12. Mean score (arbitrary units) on a coincident-timing task (15 trials/block) as a function of KR delay on two days of acquisition, and for no-KR retention tests after 10 min, 2 days, and 4 months (from Swinnen, Schmidt, Nicholson, & Shapiro, 1990).

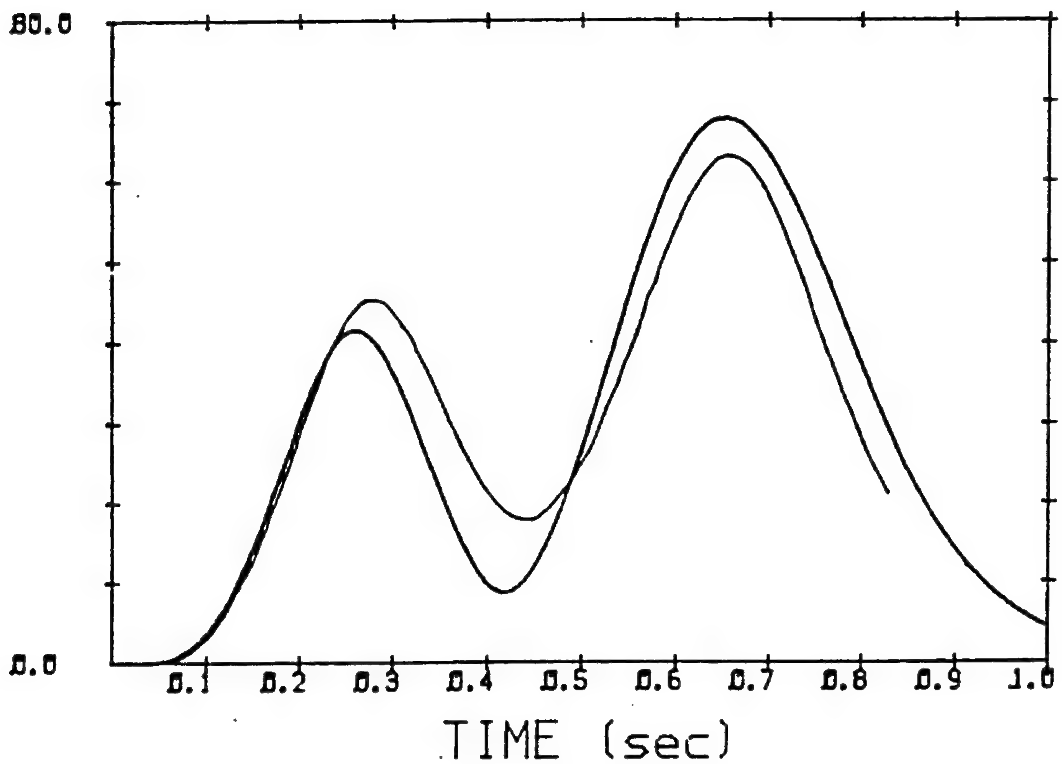


FIGURE 1

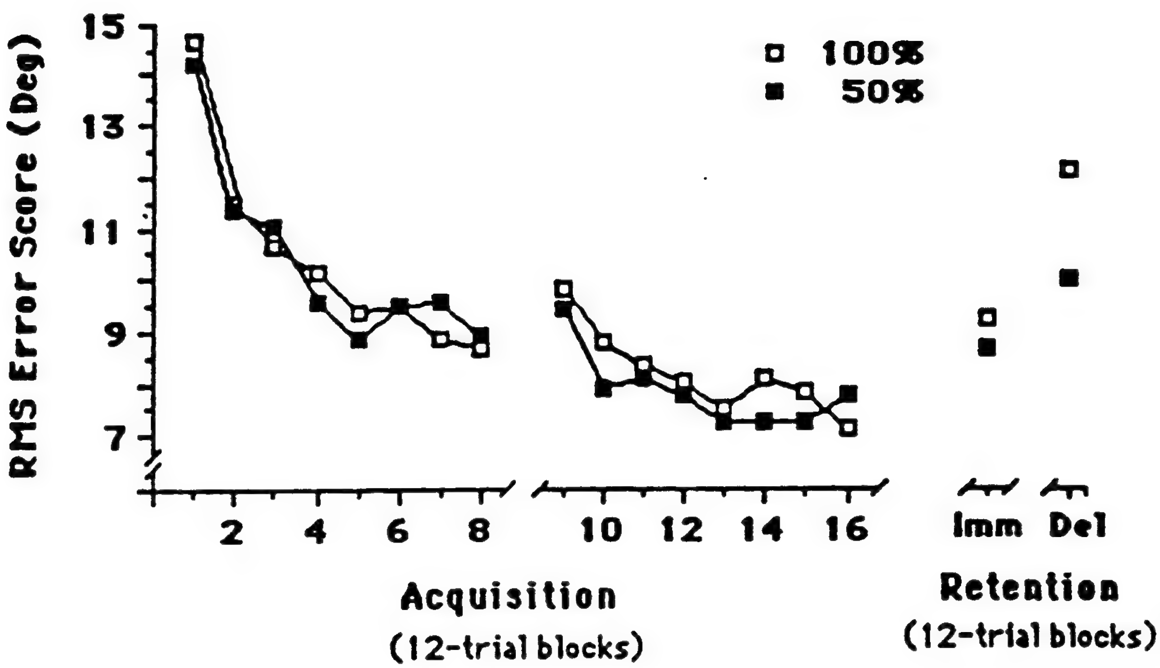
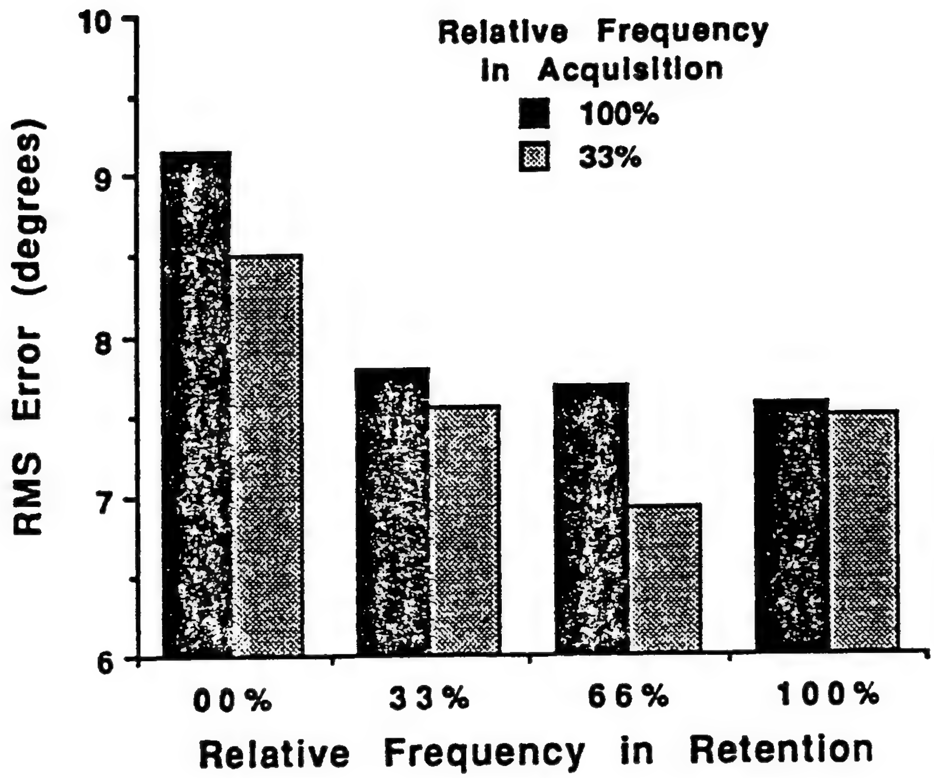
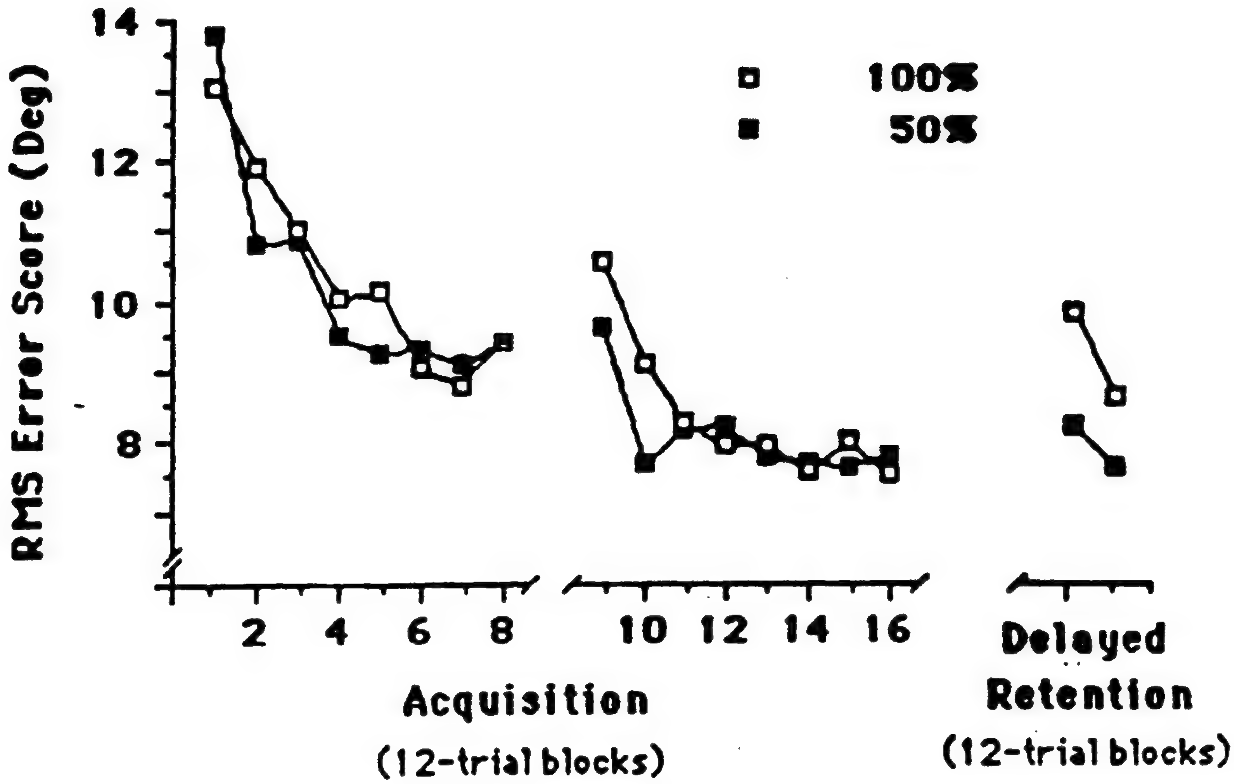

RMS Error Score (Deg)
15
13
11
9
7
100%
50%
2
4
6
8
10
12
14
16
Imm
Del
Acquisition
(12-trial blocks)
Retention
(12-trial blocks)

FIGURE 2



Relative Frequency
In Acquisition
100%
33%
RMS Error (degrees)
10
9
8
7
6
00%
33%
66%
100%
Relative Frequency in Retention

FIGURE 3



14
12
10
8
RMS Error Score (Deg)
100%
50%
2
4
6
8
10
12
14
16
Acquisition
(12-trial blocks)
Delayed
Retention
(12-trial blocks)

FIGURE 4

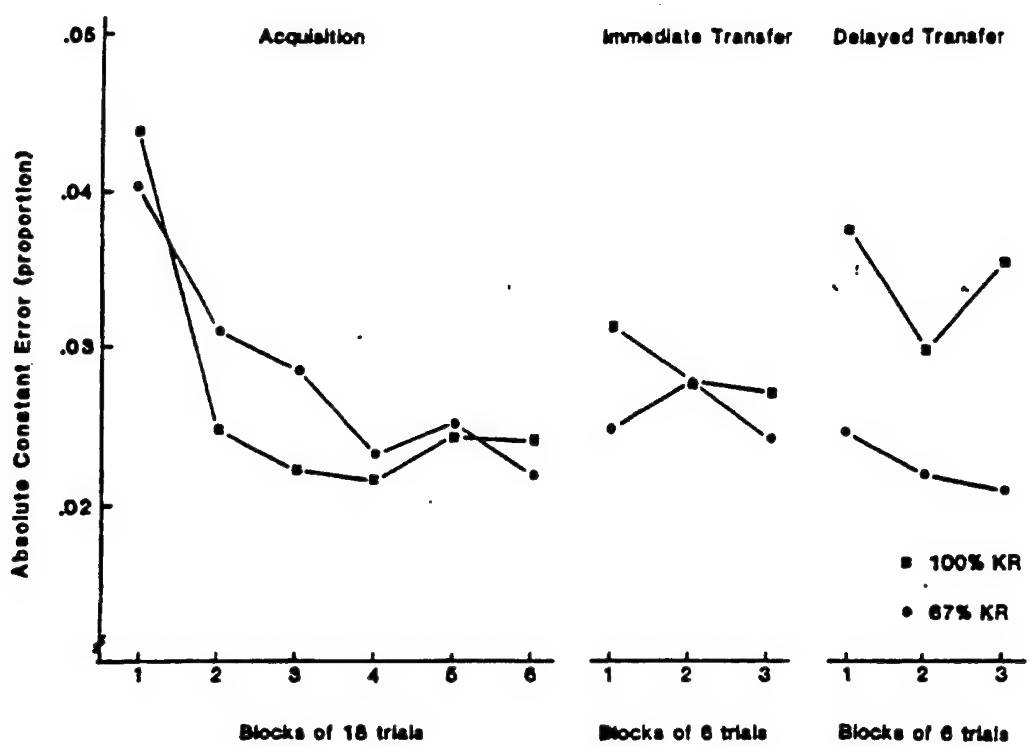

Acquisition
Immediate Transfer
Delayed Transfer
Absolute Constant Error (proportion)
.05
.04
.03
.02
100% KR
67% KR
1 2 3 4 5 6
1 2 3
1 2 3
Blocks of 18 trials
Blocks of 6 trials
Blocks of 6 trials

FIGURE 5

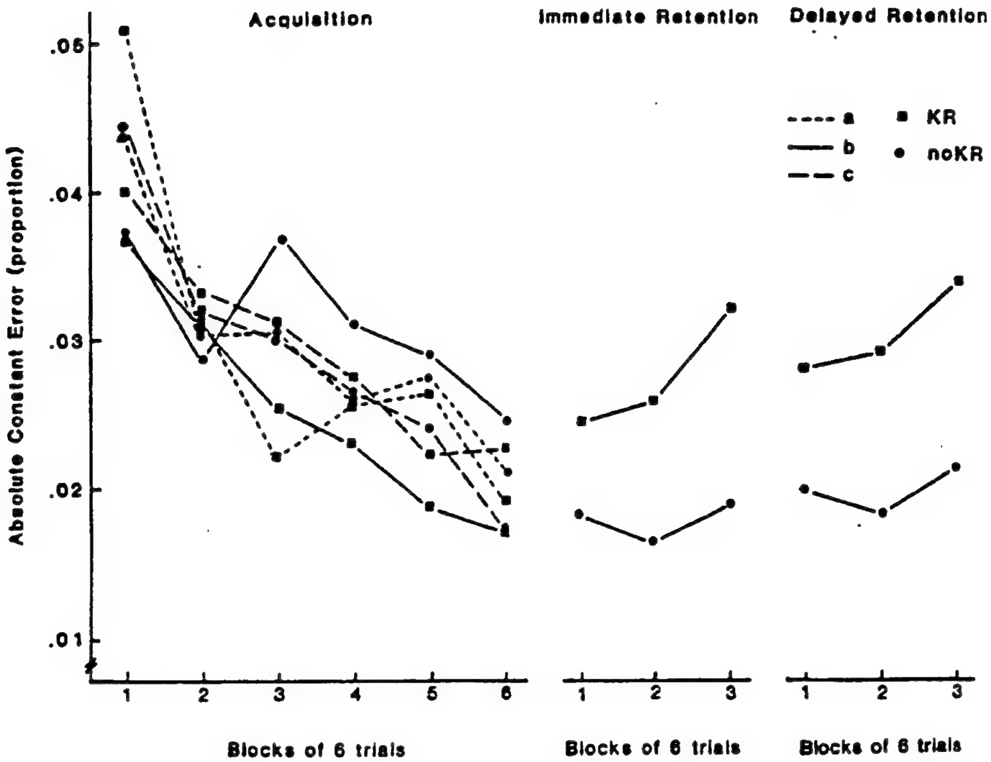

Acquisition
Immediate Retention
Delayed Retention
Absolute Constant Error (proportion)
.05
.04
.03
.02
.01
a
b
c
KR
noKR
1
2
3
4
5
6
1
2
3
1
2
3
Blocks of 6 trials
Blocks of 6 trials
Blocks of 6 trials

FIGURE 6

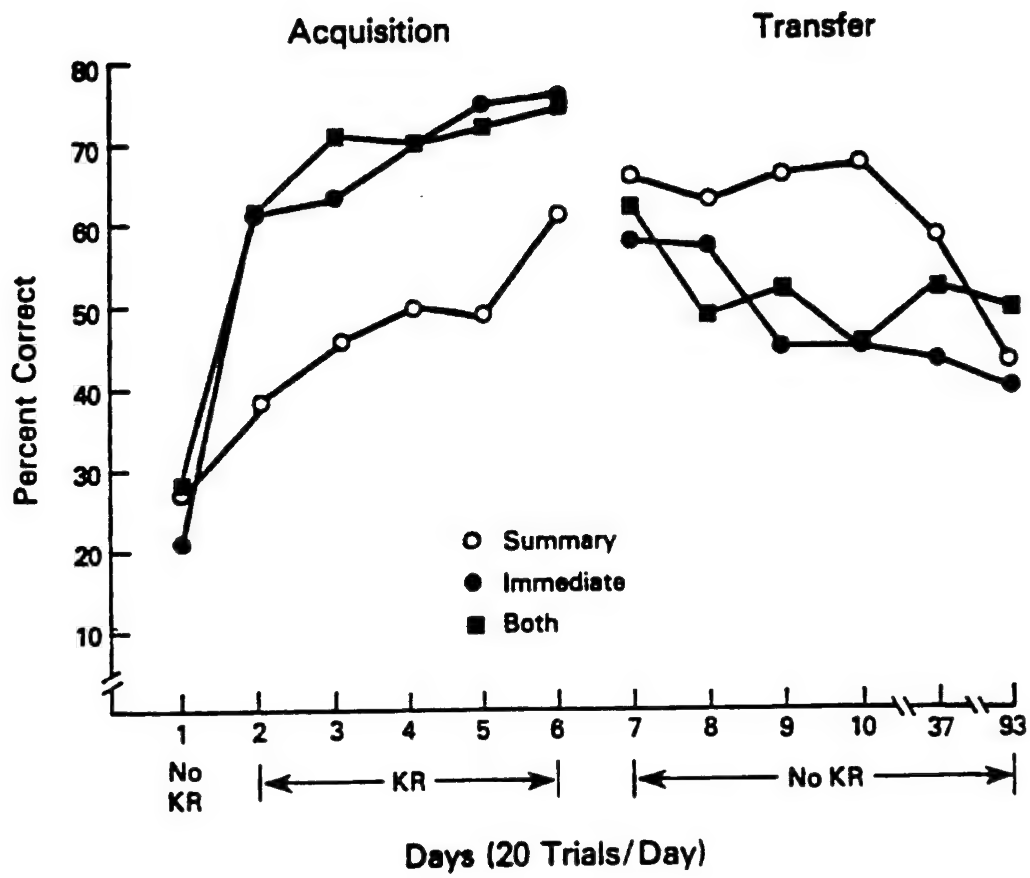

Acquisition
Transfer
Percent Correct
80
70
60
50
40
30
20
10
Summary
Immediate
Both
1
2
3
4
5
6
7
8
9
10
37
93
No KR
KR
No KR
Days (20 Trials/Day)

FIGURE 7

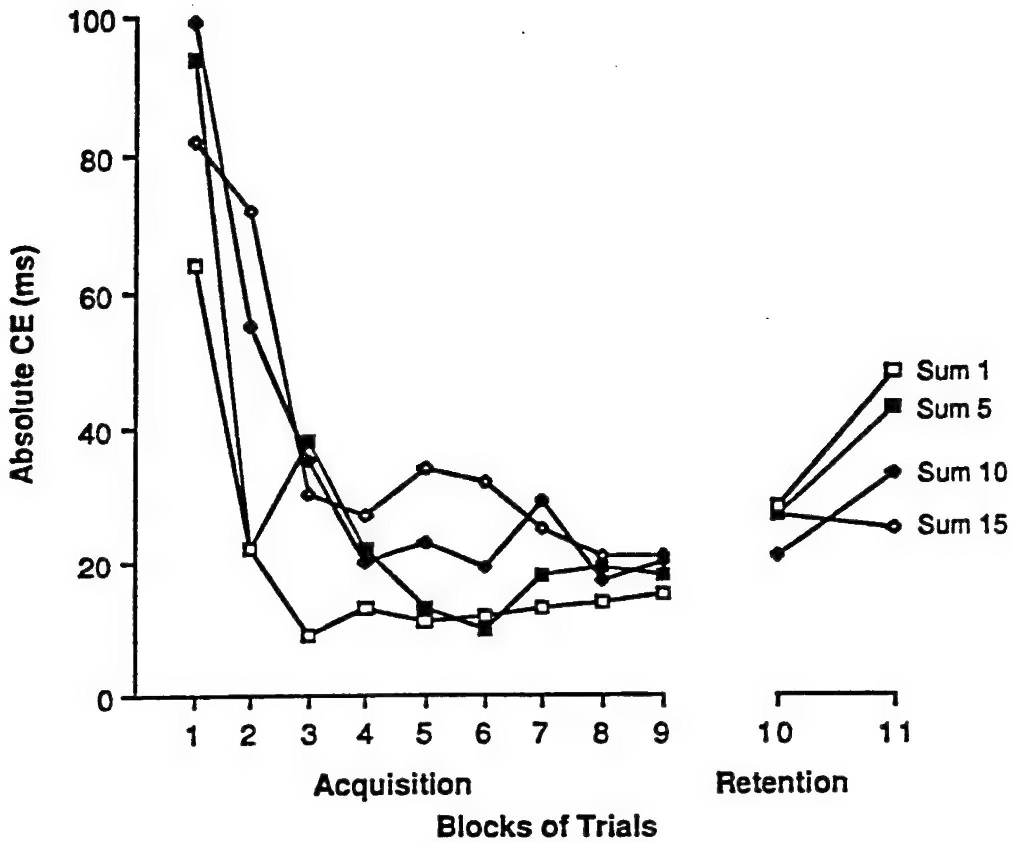

100
80
60
40
20
0
Absolute CE (ms)
1
2
3
4
5
6
7
8
9
10
11
Acquisition
Retention
Blocks of Trials
Sum 1
Sum 5
Sum 10
Sum 15

FIGURE 8

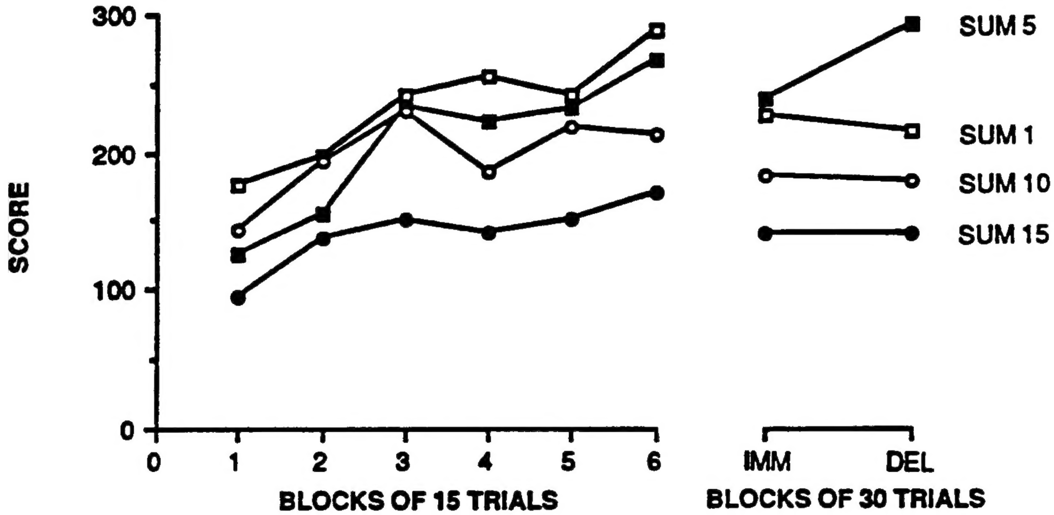

300
200
100
0
SCORE
0
1
2
3
4
5
6
BLOCKS OF 15 TRIALS
IMM
DEL
BLOCKS OF 30 TRIALS
SUM 5
SUM 1
SUM 10
SUM 15

FIGURE 9

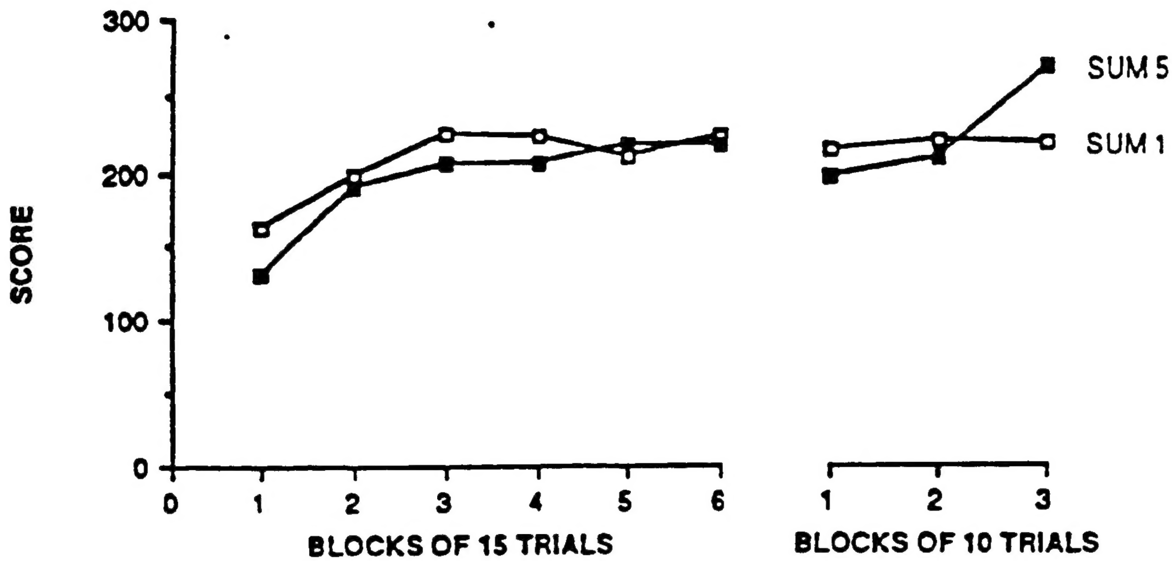

300
200
100
0
SCORE
0
1
2
3
4
5
6
1
2
3
BLOCKS OF 15 TRIALS
BLOCKS OF 10 TRIALS
SUM 5
SUM 1

FIGURE 10

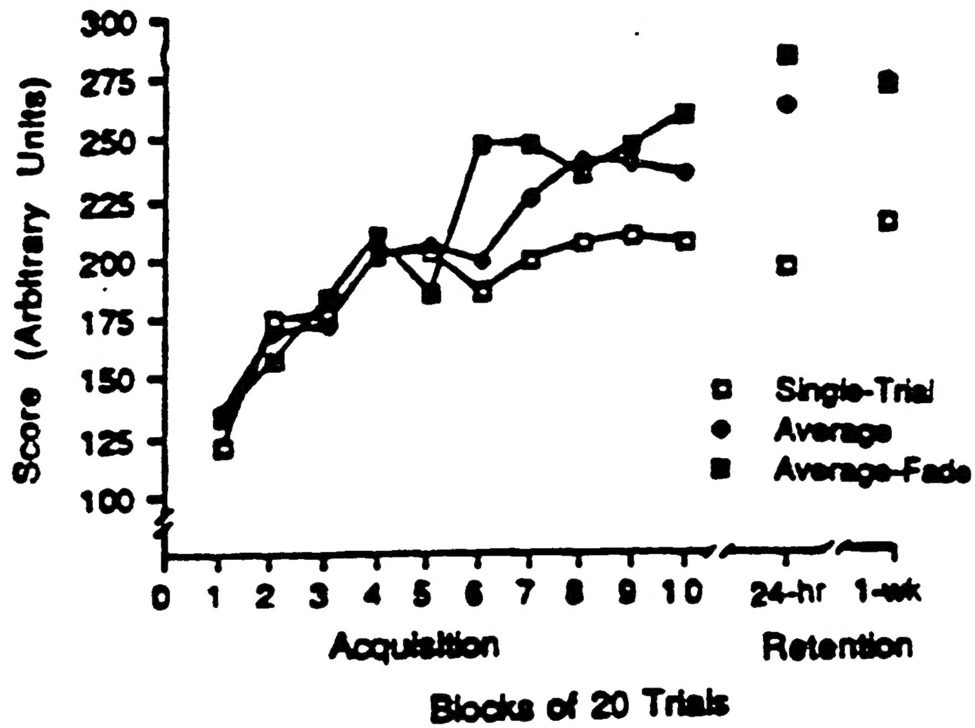

Score (Arbitrary Units)
300
275
250
225
200
175
150
125
100
Single-Trial
Average
Average-Fade
0 1 2 3 4 5 6 7 8 9 10
24-hr 1-wk
Acquisition
Retention
Blocks of 20 Trials

FIGURE 11

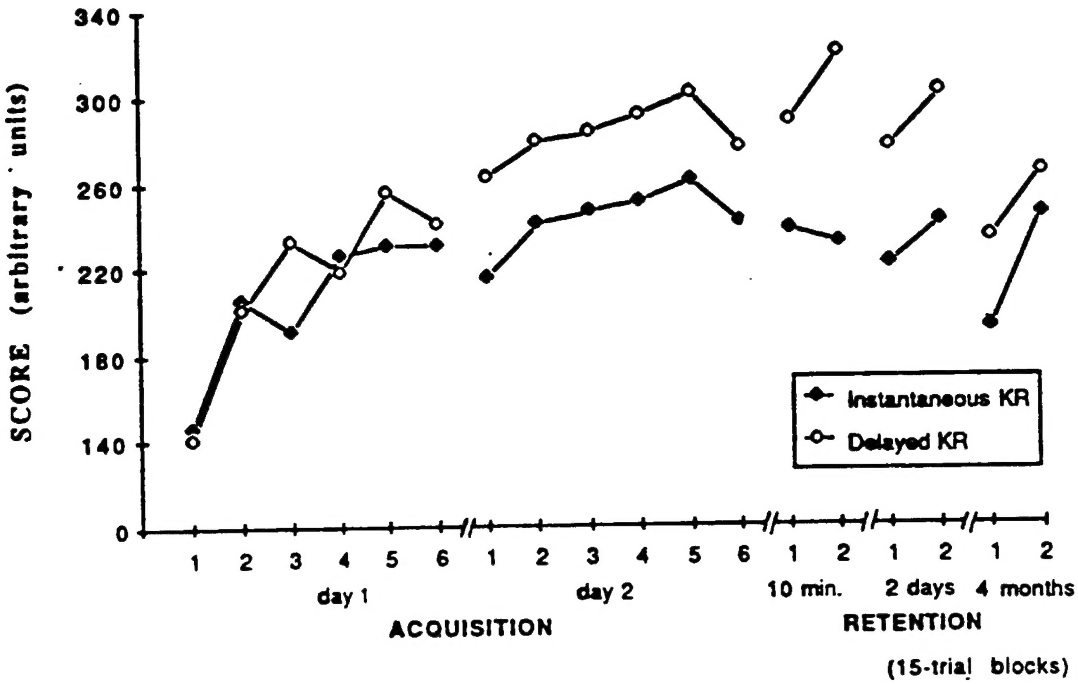

340
300
260
220
180
140
0
SCORE (arbitrary units)
1 2 3 4 5 6 1 2 3 4 5 6 1 2 1 2 1 2
day 1
day 2
10 min.
2 days
4 months
ACQUISITION
RETENTION
(15-trial blocks)
Instantaneous KR
Delayed KR

FIGURE 12